

2

# Progress with Proteome Projects: Why all Proteins Expressed by a Genome Should be Identified and How To Do It

MARC R. WILKINS<sup>1</sup>, JEAN-CHARLES SANCHEZ<sup>2</sup>, ANDREW A. GOOLEY<sup>1</sup>,  
RON D. APPEL<sup>1</sup>, IAN HUMPHERY-SMITH<sup>2</sup>, DENIS F. HOCHSTRASSER<sup>3</sup>  
AND KEITH L. WILLIAMS<sup>1,\*</sup>

<sup>1</sup> Macquarie University Centre for Analytical Biotechnology, Macquarie University, Sydney, NSW 2109, Australia; <sup>2</sup> Department of Microbiology, University of Sydney, NSW, 2006, Australia and <sup>3</sup> Central Clinical Chemistry Laboratory and Medical Computing Centre of the University of Geneva, CH 1211 Geneva 14, Switzerland

## Introduction

The advent of large genome sequencing projects has changed the scale of biology. Over a relatively short period of time, we have witnessed the elucidation of the complete nucleotide sequence for bacteriophage  $\lambda$  (Sanger *et al.*, 1982), the nucleotide sequence of an eukaryotic chromosome (Oliver *et al.*, 1992), and in the near future will see the definition of all open reading frames of some simple organisms, including *Mycoplasma pneumoniae*, *Escherichia coli*, *Saccharomyces cerevisiae*, *Caenorhabditis elegans* and *Arabidopsis thaliana*. Nevertheless, genome sequencing projects are not an end in themselves. In fact, they only represent a starting point to understanding the function of an organism. A great challenge that biologists now face is how the co-expression of thousands of genes can best be examined under physiological and pathophysiological conditions, and how these patterns of expression define an organism.

There are two approaches that can be used to examine gene expression on a large scale. One uses nucleic acid-based technology, the other protein-based technology. The most promising nucleic-acid based technology is differential display of mRNA (Liang and Pardee, 1992; Bauer *et al.*, 1993), which uses polymerase chain reaction with arbitrary primers to generate thousands of cDNA species, each which correspond to an expressed gene or part of a gene. However, it is currently unclear if this technique can be developed to reliably assay the expression of thousands of genes or

\* Corresponding Author

BEST AVAILABLE COPY

identify all cDNA species, and the approach does not easily allow a systematic screening. Analysis of gene expression by the study of proteins present in a cell or tissue presents a favorable alternative. This can be achieved by use of two-dimensional (2-D) gel electrophoresis, quantitative computer image analysis, and protein identification techniques to create 'reference maps' of all detectable proteins. Such reference maps establish patterns of normal and abnormal gene expression in the organism, and allow the examination of some post-translational protein modifications which are functionally important for many proteins. It is possible to screen proteins systematically from reference maps to establish their identities.

To define protein-based gene expression analysis, the concept of the 'proteome' was recently proposed (Wilkins *et al.*, 1995; Wasinger *et al.*, 1995). A proteome is the entire PROTEin complement expressed by a genome, or by a cell or tissue type. The concept of the proteome has some differences from that of the genome, as while there is only one definitive genome of an organism, the proteome is an entity which can change under different conditions, and can be dissimilar in different tissues of a single organism. A proteome nevertheless remains a direct product of a genome. Interestingly, the number of proteins in a proteome can exceed the number of genes present, as protein products expressed by alternative gene splicing or with different post-translational modifications are observed as separate molecules on a 2-D gel. As an extrapolation of the concept of the 'genome project', a 'proteome project' is research which seeks to identify and characterise the proteins present in a cell or tissue and define their patterns of expression.

Proteome projects present challenges of a similar magnitude to that of genome projects. Technically, the 2-D gel electrophoresis must be reproducible and of high resolution, allowing the separation and detection of the thousands of proteins in a cell. Low copy number proteins should be detectable. There should be computer gel image analysis systems that can qualitatively and quantitatively catalog the electrophoretically separated proteins, to form reference maps. A range of rapid and reliable techniques must be available for the identification and characterisation of proteins. As a consequence of a proteome project, protein databases must be assembled that contain reference information about proteins; such databases must be linked to genomic databases and protein reference maps. Databases should be widely accessible and easy to use.

Recently, there have been many changes in the techniques and resources available for the analysis of proteomes. It is the aim of this chapter to discuss the status of the areas outlined above, and to review briefly the progress of some current proteome projects.

## Two-dimensional electrophoresis of proteomes

Two dimensional (2-D) gel electrophoresis involves the separation of proteins by their isoelectric point in the first dimension, then separation according to molecular weight by sodium dodecyl sulfate electrophoresis in the second dimension. Since first described (Klose, 1975; O'Farrell, 1975; Scheele, 1975), it has become the method of choice for the separation of complex mixtures of proteins, albeit with many modifications to the original techniques. 2-D electrophoresis forms the basis of proteome projects through separating proteins by their size and charge (Hochstrasser *et al.*,

Fig-  
illu-  
was  
ing  
the  
of L

19  
pr  
sin

2-1

A;  
ph-  
en

## HEPG2 2D-PAGE MAP

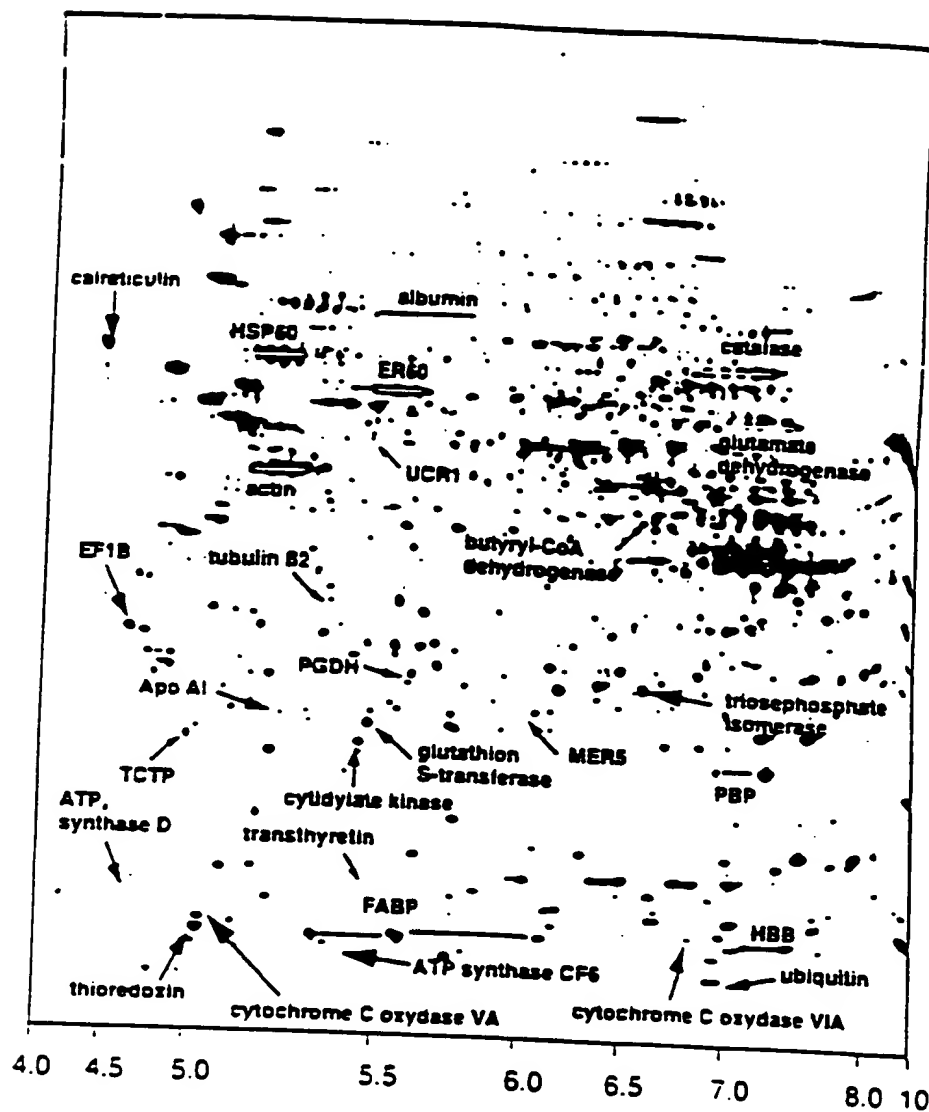


Figure 1. Two-dimensional gel electrophoresis map of a human hepatoblastoma-derived cell line, illustrating the very high resolution of the technique. The first dimensional separation (right to left of figure) was achieved using immobilised pH gradient electrophoresis of 4.0 to 10.0 units. The second dimension (top to bottom of figure) was SDS-PAGE using a 11%–14% acrylamide gradient, allowing separation in the molecular weight range 10–250 kDa. Proteins were visualised by silver staining. Arrows show proteins of known identity.

1992; Celis *et al.*, 1993; Garrels and Franza, 1989; VanBogelen *et al.*, 1992). Current protocols can resolve two to three thousand proteins from a complex sample on a single gel (Figure 1).

#### 2-D GEL RESOLUTION AND REPRODUCIBILITY

A primary challenge of separating complex mixtures of proteins by 2-D gel electrophoresis has been to achieve high resolution and reproducibility. High resolution ensures that a maximum of protein species are separated, and high reproducibility is

vital to allow comparison of gels from day to day and between research sites. These factors can be difficult to achieve.

Carrier ampholytes are a common means of isoelectric focusing for the first dimension of 2-D electrophoresis. Gels are usually focused to equilibrium to separate proteins in the pI range 4 to 8, and run in a non-equilibrium mode (NEPHGE) to separate proteins of higher pI (7 to 11.5) (O'Farrell, 1975; O'Farrell, Goodman and O'Farrell, 1977). Unfortunately, the use of carrier ampholytes in the isoelectric focusing procedure is susceptible to 'cathode drift', whereby pH gradients established by prefocusing of ampholytes slowly change with time (Righetti and Drysdale, 1973). Carrier ampholyte pH gradients are also distorted by high salt concentration of samples (Bjellqvist *et al.*, 1982), and by high protein load (O'Farrell, 1975). A further limitation is that isoelectric focusing gels, which are cast and subject to electrophoresis in narrow glass tubes, need to be extruded by mechanical means before application to the second dimension – a procedure that potentially distorts the gel. Nevertheless, many of the above shortcomings can be avoided by loading small amounts of  $^{14}\text{C}$  or  $^{35}\text{S}$  radiolabelled samples (Garrels, 1989; Neidhardt *et al.*, 1989; Vandekerckhove *et al.*, 1990). High sensitivity detection is then achieved through use of fluorography or phosphorimaging plates (Bonner and Laskey, 1974; Johnston, Pickett and Barker, 1990; Patterson and Lutter, 1993). However, this approach is only practicable for organisms or tissues that can be radiolabelled.

An alternative technique, which is becoming the method of choice for the first dimension separation of proteins, involves isoelectric focusing in immobilized pH gradient (IPG) gels (Bjellqvist *et al.*, 1982; Görg, Postel and Gunther, 1988; Righetti, 1990). Immobilized pH gradients are formed by the covalent coupling of the pH gradient into an acrylamide matrix, creating a gradient that is completely stable with time. IPG gels are usually poured onto a stiff backing film, which is mechanically strong and provides easy gel handling (Ostergren, Eriksson and Bjellqvist, 1988). The major advantages of IPG separations are that they do not suffer from cathodic drift, they allow focusing of basic and very acidic proteins to equilibrium, pH gradients can be precisely tailored (linear, stepwise, sigmoidal), and that separations over a very narrow pH range are possible (0.05 pH units per cm) (Righetti, 1990; Bjellqvist *et al.*, 1982, 1993a; Sinha *et al.*, 1990; Görg *et al.*, 1988; Gelfi *et al.*, 1987; Gunther *et al.*, 1988). However, it is not currently possible to use IPG gels to separate very basic proteins of isoelectric point greater than 10, although this is under development. Narrow pH range separations are useful to address problems of protein co-migration in complex samples, allowing 'zooming in' on regions of a gel (Figure 2). IPG gel strips are now commercially available, which begin to address the problems of intra- and inter-lab isoelectric focusing reproducibility.

There are two means of electrophoresis for the second dimension separation of proteins: vertical slab gels and horizontal ultrathin gels (Görg, Postel, and Gunther, 1988). Both are usually SDS-containing gradient gels of approximately 11% to 15% acrylamide, which separate proteins in the molecular mass range of 10 – 150 kD. A stacking gel is not usually used with slab gels, but is necessary when using horizontal gel setups (Görg, Postel and Gunther, 1988). Comparisons have shown that there is little or no difference in the reproducibility of electrophoresis using either approach (Corbett *et al.*, 1994a), but commercially available vertical or horizontal precast gels will provide greater reproducibility for occasional users. For slab gel electrophoresis,

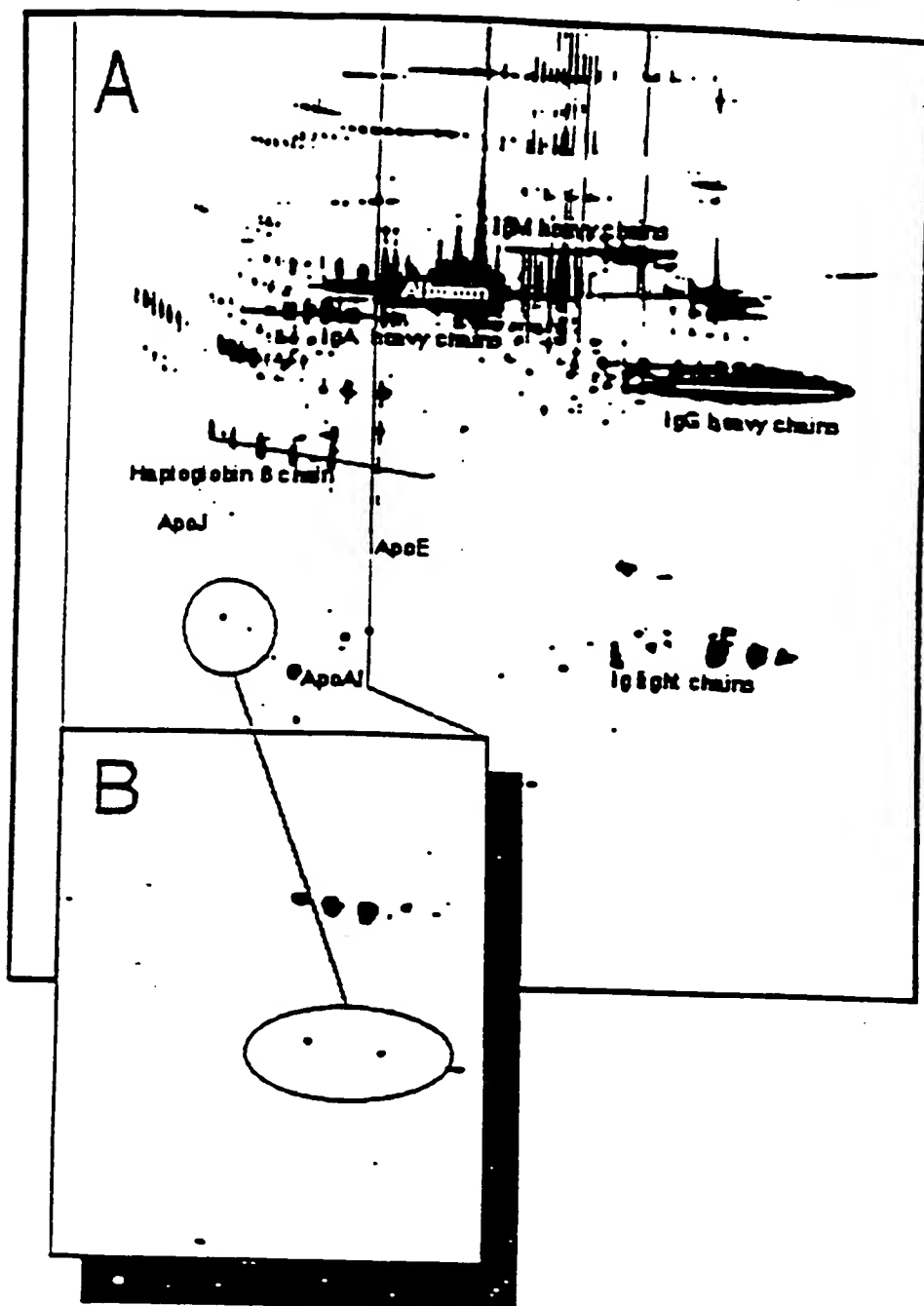


Figure 2. Two-dimensional gel electrophoresis allows 'zooming in' on areas of interest. Rings highlight 2 proteins common to each gel. (A) Wide pI range two dimensional electrophoresis map of human plasma proteins. First dimension separation was achieved using an immobilised pH gradient of 3.5 to 10.0 units. The second dimension was SDS-PAGE. Actual gel size was 16cm x 20cm, and proteins were visualised with silver staining. (B) Narrow pI range electrophoresis was used to 'zoom in' on a small region of the plasma map. The first dimension used a narrow range immobilised pH gradient of 4.2 to 5.2 units, and second dimension was SDS-PAGE. Micropreparative loading was used, and the gel blotted to PVDF. Proteins were visualised with amido black. Actual blot size was 16cm x 20cm.

the use of piperazine diacrylyl as a gel crosslinker and the addition of thiosulfate in the catalyst system has been shown to give better resolution and higher sensitivity detection (Hochstrasser and Merril, 1988; Hochstrasser, Patchornik and Merril, 1988).

Notwithstanding the advances described above, there is an increasing demand to improve the reproducibility of 2-D electrophoresis to facilitate database construction and proteome studies. Harrington *et al.* (1993) explain that if a gel resolves 4000 protein spots, and there is 99.5% spot matching from gel to gel, this will produce 20 spot errors per gel. This amount of error, which might accumulate with each gel to gel comparison used in database construction, could produce an unacceptable degree of uncertainty in gel databases. To address these issues, partial automation of large 2-D gel separations has been undertaken (Nokihara, Morita and Kuriki, 1992; Harrington *et al.*, 1993). Although results are preliminary, spot to spot positional reproducibility in one study was found to be threefold improved over manual methods (Harrington *et al.*, 1993). It should be noted that small 2-D gel formats (50 × 43 mm) have been almost completely automated (Brewer *et al.*, 1986), although these are not generally used for database studies.

#### MICROPREPARATIVE 2-D GEL ELECTROPHORESIS

With the advent of affordable protein microcharacterisation techniques, including N-terminal microsequencing, amino acid analysis, peptide mass fingerprinting, phosphate analysis and monosaccharide compositional analysis, a new challenge for 2-D electrophoresis has been to maintain high resolution and reproducibility but to provide protein in sufficient quantities for chemical analysis (high nanogram to low microgram quantities of proteins per spot). This becomes difficult to achieve with very complex samples such as whole bacterial cells, as the initial protein load is divided among 2000 to 4000 protein species. Two approaches are used for producing amounts of material that can be chemically characterised. The first method is to run multiple gels, collect and pool the spots of interest, and subject them to concentration (Ji *et al.*, 1994; Walsh *et al.*, 1995; Rasmussen *et al.*, 1992). In this approach, the concentration process must also act as a purification step to remove accumulated electrophoretic contaminants such as glycine. A more elegant approach has been to exploit the high loading capacity of IPG isoelectric focusing. The high loading capacity of immobilised pH gradients was described early (Ek, Bjellqvist and Righetti, 1983), but has only recently been applied to 2-D electrophoresis (Hanash *et al.*, 1991; Bjellqvist *et al.*, 1993b). Up to 15 mg of protein can be applied to a single gel, yielding microgram quantities of hundreds of protein species. A further benefit of this approach is that proteins present in low abundance, which may not be visualised by lower protein loads, are more likely to be detected. The use of electrophoretic or chromatographic prefractionation techniques (Hochstrasser *et al.*, 1991a; Harrington *et al.*, 1992), followed by high loading of narrow-range IPG separations (Bjellqvist *et al.*, 1993b) provides a likely solution to studies on proteins present in low abundance.

#### Methods of protein detection

There are many means for detecting proteins from 2-D gels. The method used will be dictated by factors including protein load on gel (analytical or preparative), the purpose of the gel (for protein quantitation or for blotting and chemical characterisation), and the sensitivity required. The most common means of protein detection and their applications are shown in *Table 1*. Most detection methods have drawbacks, for

Table 1: Common stains for 2-D gels or blots and their applications.

Detection Method	Main applications	Unsuitable applications	Sensitivity	References
[ <sup>35</sup> S] Met or <sup>14</sup> C radiolabelling and fluorography or phosphorimaging	Cell lines, cultured organisms	Samples that cannot be labelled	20 ppm of radiolabel in a spot	Garcia and Franza, 1990 Latham, Garret and Siller, 1993
[ <sup>125</sup> I]thiourea silver	Extremely high sensitivity gel staining	Preparative 2-D, PVDF or NC membranes	0.4 ng protein on spot or band of gel	Wallace and Saluz, 1992a,b
Silver	Very high sensitivity gel staining, can be mono or polychromatic	Preparative 2-D, PVDF or NC membranes	4 ng protein on spot or band of gel	Rahilloud, 1992 Hochstrasser and Merril, 1988
Coomassie blue R-250	Staining of gels, staining of PVDF membranes before protein sequencing	Staining prior to direct mass determination from PVDF; amino acid analysis on PVDF; detection of some glycoproteins	40 ng protein on band or spot of gel	Sirupat <i>et al.</i> , 1994; Gharahdaghi <i>et al.</i> , 1992; Goldberg <i>et al.</i> , 1988; Sanchez <i>et al.</i> , 1992
Colloidal gold	Staining NC membranes, staining PVDF before direct MALDI-TOF	Gels	60 x higher than coomassie	Yamaguchi and Asakawa, 1988; Eckerskorn <i>et al.</i> , 1992; Sirupat <i>et al.</i> , 1994
Zinc imidazole	Reverse staining of gels or membranes; may be beneficial in MALDI-TOF of peptides	Where positive image is required	Higher than coomassie	Ortiz <i>et al.</i> , 1992; James <i>et al.</i> , 1993
Ponceau S and amido black	Staining higher protein loads on PVDF, for protein sequencing or amino acid analysis	Staining prior to direct mass determination from PVDF	100 ng protein on band or spot of gel	Sanchez <i>et al.</i> , 1992; Sirupat <i>et al.</i> , 1994; Wilkins <i>et al.</i> , 1995
India ink	Staining of membrane-bound proteins, staining PVDF before direct MALDI-TOF	Gel staining, not quantitative from protein to protein	1-10 ng	Li <i>et al.</i> , 1989; Hughes, Mack and Hamparian, 1988; Sirupat <i>et al.</i> , 1994
Stains-all	Staining to detect glycoproteins or Ca <sup>2+</sup> binding proteins	General gel staining	100 ng protein on band or spot of gel	Campbell, MacLennan and Jorgensen, 1983; Goldberg <i>et al.</i> , 1988

PVDF = polyvinylidene difluoride, NC = nitrocellulose, MALDI-TOF = matrix assisted laser desorption/ionisation time of flight mass spectrometry.

example, some glycoproteins are not stained by coomassie blue (Goldberg *et al.*, 1988), and many organic dyes are unsuitable for protein detection on PVDF if samples are to be used for direct matrix-assisted laser desorption/ionisation mass spectrometry (Sirupat *et al.*, 1994).

Although most means of protein detection give some indication of the quantities of protein present, in general they cannot be used for global quantitation. This is because

no protein, stain is able consistently to detect proteins over a wide range of concentrations, isoelectric points and amino acid compositions, and with a variety of post-translational modifications (Goldberg *et al.*, 1988; Li *et al.*, 1989). Furthermore, there are large differences in staining pattern when identical gels or blots are subjected to different stains, including amido black, imidazole zinc, india ink, ponceau S, colloidal gold, or coomassie blue (Tovey, Ford and Baldo, 1987; Ortiz *et al.*, 1992). The most common means of quantitating large numbers of proteins in a 2-D gel involves the radiolabelling of protein samples prior to electrophoresis, and protein quantitation based on fluorography and image analysis or liquid scintillation counting (Garrele, 1989; Celis and Olsen, 1994). However, proteins which do not contain methionine cannot be detected if only [<sup>35</sup>S] methionine is used for labelling. Amino acid analysis of protein spots visualised by other techniques presents a likely means of protein quantitation for the future.

#### BLOTTING OF PROTEINS TO MEMBRANES

Electrophoretic blotting of proteins from two-dimensional polyacrylamide gels to membranes presents many options for protein identification and microcharacterisation which are not possible when proteins remain in gels. For example, when proteins are blotted to polyvinylidene difluoride (PVDF) membranes, they can be identified by N-terminal sequencing, amino acid analysis, or immunoblotting, or they may be subjected to endoproteinase digestion, monosaccharide analysis, phosphate analysis, or direct matrix-assisted laser desorption ionisation mass spectrometry (Matsudaira, 1987; Wilkins *et al.*, 1995; Jungblut *et al.*, 1994; Sutton *et al.*, 1995; Rasmussen *et al.*, 1994; Weizthandler *et al.*, 1993; Murthy and Iqbal, 1991; Eckerskorn *et al.*, 1992). It is possible to combine some of these procedures on a single protein spot on a PVDF membrane (Packer *et al.*, 1995; Wilkins *et al.*, submitted; Weizthandler *et al.*, 1993). This is useful when minimal amounts of protein are available for analysis. These techniques will be explored in detail later in this review. Notwithstanding the above, there are some disadvantages associated with blotting of proteins to membranes. There is always loss of sample during blotting procedures (Eckerskorn and Lottspeich, 1993), and common protein detection methods are less sensitive or not applicable to membranes (Table 1), presenting difficulties for the analysis of low abundance proteins. Detailed discussion of the merits of available membranes and common blotting techniques can be found elsewhere (Eckerskorn and Lottspeich, 1993; Strupat *et al.*, 1994; Patterson, 1994).

#### 2-D gel analysis, documentation, and proteome databases

Following protein electrophoresis and detection, detailed analysis of gel images is undertaken with computer systems. For proteome projects, the aim of this analysis is to catalogue all spots from the 2-D gel in a qualitative and if possible quantitative manner, so as to define the number of proteins present and their levels of expression. Reference gel images, constructed from one or more gels, form the basis of two-dimensional gel databases. These databases also contain protein spot identities and



details of their post-translational modifications. 2-D gel databases are beginning to be linked to or integrated with comprehensive protein and nucleic acid databases (Neidhardt *et al.*, 1989; Simpson *et al.*, 1992; Appel *et al.*, 1994), and 'organism' databases, containing DNA sequence data, chromosomal map locations, reference 2-D gels and protein functional information for an organism, are becoming established as genome and proteome projects progress (VanBogelen *et al.*, 1992; Yeast Protein Database cited in Garrels *et al.*, 1994).

## GEL IMAGE ANALYSIS AND REFERENCE GELS

After 2-D electrophoresis and protein visualisation by staining, fluorography or phosphorimaging, images of gels are digitised for computer analysis by an image scanner, laser densitometer, or charge-coupled device (CCD) camera (Garrels, 1989; Celis *et al.*, 1990a; Urwin and Jackson, 1993). All systems digitise gels with a resolution of 100 – 200 mm, and can detect a wide range of densities or shading (256 or more 'grey scales'). Following this, gel images are subjected to a series of manipulations to remove vertical and horizontal streaking and background haze, to detect spot positions and boundaries, and to calculate spot intensity (Figure 3). A standard spot (SSP) number, containing vertical and horizontal positional information, is assigned to each detected spot and becomes the protein's reference number. Table 2 lists some notable software packages which process 2-D gel images.

Table 2: Some Software Packages for the Analysis of Gel Images.

Gel Image Analysis System	References*
ELSIE I & II	Olsen and Miller, 1988; Wirth <i>et al.</i> , 1991; Wirth <i>et al.</i> , 1993
GELLAB I & II	Wu, Lemkin and Upton, 1993; Lemkin, Wu and Upton, 1993; Myrick <i>et al.</i> , 1993
MELANIE I & II	Appel, <i>et al.</i> 1991; Hochstrasser <i>et al.</i> 1991b
QUEST I & II and PDQUEST	Garrels, 1989; Monardo <i>et al.</i> , 1992; Holt <i>et al.</i> , 1992; Celis <i>et al.</i> , 1990a,b
TYCHO & KEPLAR	Anderson <i>et al.</i> , 1982; Richardson, Horn and Anderson, 1993

\* These references are not exhaustive; they include some references of use as well as authors of the system

As there are difficulties in the electrophoresis of samples with 100% reproducibility, reference gel images are often constructed from many gels of the same sample (Garrels and Franza, 1989; Neidhardt *et al.*, 1989). Since this involves the matching of 2000 to 4000 proteins from one gel to another, it presents a considerable challenge to image analysis systems. Matching of gels is usually initiated by an operator, who manually designates approximately 50 or so prominent spots as 'landmarks' on gels to be cross-matched. Proteins which match are then established around landmarks, using computer-based vector algorithms to extend the matching over the entire gel. Close to 100% of spots from complex samples can be matched by these methods, although different degrees of operator intervention may be required (Olsen and Miller, 1988; Lemkin and Lester, 1989; Garrels, 1989; Myrick *et al.*, 1993).

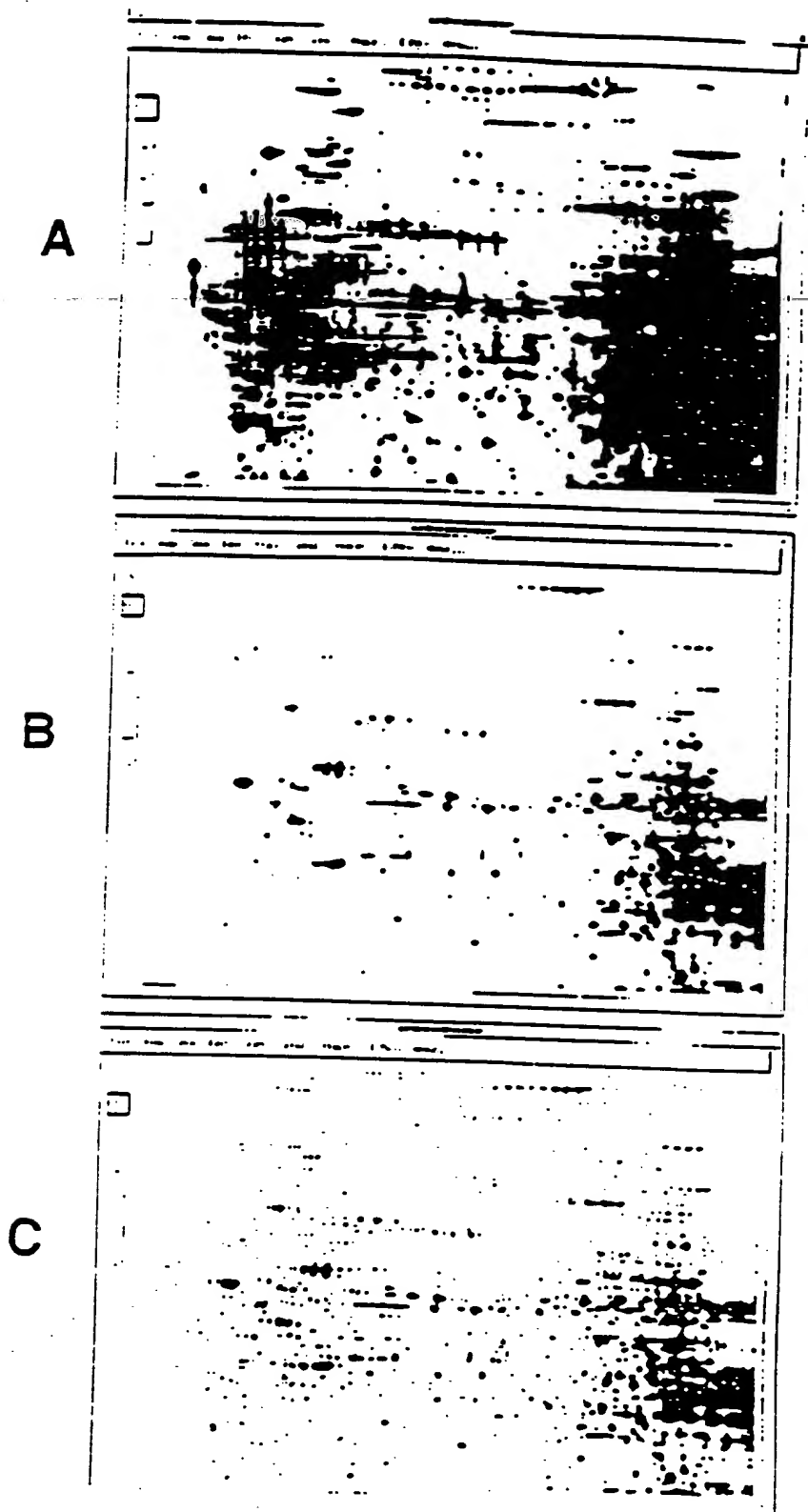


Figure 3. Computer processing of gel images. Shown is a wide pI range 2-D separation of human liver proteins, processed by Melanie software (Appel *et al.*, 1991). (A) Original gel image as captured by laser densitometer. (B) Gel image after processing to remove streaking and background. (C) Outline definition of all spots on the gel.

By  
2-D  
during  
are rel  
obtain  
curves  
MWL  
Bogel  
*et al.*, 19  
to PVI  
*et al.*  
protein  
amino  
carry  
positi

SPOT C

A maj  
separa  
to dete  
positi  
of ma  
perform  
minute  
scintill  
by rel  
protein  
*et al.*,  
Garrel  
be no  
Limit  
not ac  
only e  
an alt  
Myric  
1992.

Wh  
and m  
protei  
their  
transf  
regul  
synth  
(Lath.

## CALCULATION OF PROTEIN ISOELECTRIC POINT AND MOLECULAR WEIGHT

Estimation of the isoelectric point (pI) and molecular weight (MW) of proteins from 2-D gels provides fundamental parameters for each protein, which are also of use during identification procedures (see following section). The pI and MW of proteins are recorded in 2-D gel databases. Accurate estimations of protein pI and MW can be obtained by using 20 or more known proteins on a reference map to construct standard curves of pI and molecular weight, which are then used to calculate estimated pI and MW of unknown proteins (Neidhardt *et al.*, 1989; Garrels and Franza, 1989; Van-Bogelen, Hutton and Neidhardt, 1990; Anderson and Anderson, 1991; Anderson *et al.*, 1991; Latham *et al.*, 1992). Alternatively, the MW of individual proteins blotted to PVDF can be determined very accurately by direct mass spectrometry (Eckerskorn *et al.*, 1992). Where immobilised pH gradients are used, the focusing position of proteins allows their pI to be measured within 0.15 units of that calculated from the amino acid sequence (Bjellqvist *et al.*, 1993c). It must be noted, however, that proteins carrying post-translational modifications may migrate to unexpected pI or MW positions during electrophoresis (Packer *et al.*, 1995).

## SPOT QUANTITATION AND EXPRESSION ANALYSIS

A major challenge faced in proteomic projects is the quantitative analysis of proteins separated by 2-D electrophoresis. The most accurate means of protein quantitation is to determine chemically the amount of each protein present by amino acid compositional analysis. However, the current method of choice for quantitative analysis of many proteins is to radiolabel samples with [<sup>35</sup>S] methionine or <sup>14</sup>C amino acids, perform the 2-D electrophoresis, and measure protein levels in disintegrations per minute (dpm) or units of optical density. Quantitation is achieved either by liquid scintillation counting, or by gel image analysis where spot densities are quantitated by reference to gel calibration strips containing known amounts of radiolabelled protein or against the integrated optical density of all spots visualised (Vandekerckhove *et al.*, 1990; Celis *et al.*, 1990b; Celis and Olsen, 1994; Garrels, 1989; Latham, Garrels and Solter, 1993; Fey *et al.*, 1994). All approaches effectively allow spots to be normalised against the total disintegrations per minute loaded onto the gel. Limitations that remain with radiolabelling methods are that absolute quantitation is not achieved because all proteins have varying amounts of any amino acid, and that only easily labelled samples can be investigated. Quantitative silver staining presents an alternative (Giometti *et al.*, 1991; Harrington *et al.*, 1992; Rodriguez *et al.*, 1993; Myrick *et al.*, 1993), which when undertaken with [<sup>35</sup>S]thiourea (Wallace and Saluz, 1992 a,b) is of extremely high sensitivity.

When protein spots from samples prepared under different conditions are quantitated and matched from gel to gel, it becomes possible to examine changes and patterns in protein expression. Large scale investigation of up- and down-regulation of proteins, their appearance and disappearance, can be undertaken. For example, simian virus 40 transformed human keratinocytes were shown to have 177 up-regulated and 58 down-regulated proteins compared to normal keratinocytes (Celis and Olsen, 1994); detailed synthesis profiles of 1200 proteins have been established in 1 to 4 cell mouse embryos (Latham *et al.*, 1991, 1992); and 4 proteins out of 1971 were found to be markers for

cadmium toxicity in urinary proteins (Myrick *et al.*, 1993). Complex global changes in protein expression as a result of gene disruptions have also been investigated (S. Fey and P. Most-Larsen, Personal communication). Impressively, large gel sets showing protein expression under different conditions can be globally investigated using statistical methods that find groups of related objects within a set. For example, the REF52 rat cell line database, consisting of 79 gels from 12 experimental groups where each gel contains quantitative data for 1600 cross-matched proteins, has been analysed by cluster analysis (Garrels *et al.*, 1990). This revealed clusters of proteins that, for example, were induced or repressed similarly under simian virus 40 or adenovirus transformation, suggesting a common mechanism. Protein groups that were induced or repressed during culture growth to confluence were also found. It is obvious that the potential for investigation of cellular control mechanisms by these approaches is immense. It is equally clear that investigations of gene expression of this scale are currently technically impossible using nucleic-acid based techniques.

Table 3: Some proteomic databases and their special features

Proteomic database	Special features	References
<i>E. coli</i> gene-protein database	Gel spots linked with GenBank and Kohara clones; quantitative spot measurements under different growth conditions	VanBogelen and Neidhardt, 1991. VanBogelen <i>et al.</i> , 1992
Human heart databases	Identification of disease markers; two separate databases have been established	Baker <i>et al.</i> , 1992 Cornett <i>et al.</i> , 1993b Junghans <i>et al.</i> , 1993
Human keratinocyte database	Extensive identifications; quantitative spot measurements of transformed cells; identification of disease markers	Celis <i>et al.</i> , 1990a Celis <i>et al.</i> , 1993 Celis and Olsen, 1993
Mouse embryo database	Quantitative spot measurements through 1 to 4 cell stage	Latham <i>et al.</i> , 1991 Latham <i>et al.</i> , 1992
Mouse liver database (Argonne Protein Mapping Group)	Documents changes due to exposure to ionizing radiation and toxic chemicals	Gionetti, Taylor and Tollaksen, 1992
Rat liver epithelial database	Detailed subcellular fractionation studies	Wirth <i>et al.</i> , 1991 Wirth <i>et al.</i> , 1993
Rat liver database	Extensive studies on regulation of proteins by drugs and toxic agents	Anderson and Anderson, 1991. Anderson <i>et al.</i> , 1992. Richardson, Horn and Anderson, 1993
REF 52 rat cell line database	Accessible via World Wide Web; quantitative spot measurements under different conditions	Garrels and Franza 1989 Boutell <i>et al.</i> , 1993
SWISS-2DPAGE containing human reference maps	Accessible via World Wide Web; completely integrated with SWISS-PROT and SWISS-3DIMAGE	Appel <i>et al.</i> , 1993 Hochstrasser <i>et al.</i> , 1992 Hughes <i>et al.</i> , 1993 Golaz <i>et al.</i> , 1993
Yeast Protein Database (YPD) and Yeast Electrophoretic Protein Database (YEPD)	Completely crossreferenced organism database. YPD has extensive information on over 2500 proteins; YEPD has many identifications	Garrels <i>et al.</i> , 1993

Protein  
protein  
informa  
2-D ge  
subcell  
of rete  
should  
Macint  
the are  
annota  
sequen  
One  
SWISS  
1993;  
feature  
2DPA

Table 4  
All thre  
expans  
Inform

Annua

Cross-  
Refer  
Data

Other

## FEATURES OF PROTEOME DATABASES

Proteome projects rely heavily on computer databases to store information about all proteins expressed by an organism. 'Proteome databases' should contain detailed information of proteins already characterised elsewhere, as well as protein data from 2-D gels such as apparent pI and MW, expression level under different conditions, subcellular localisation, and information on post-translational modifications. Images of reference 2-D gels, showing protein SSP numbers and protein identifications, should also be included. Ideally, proteome databases should be accessible with Macintosh or IBM personal computers and easy to use. Some proteome databases and the areas they cover are listed in *Table 3*. Databases range from collections of unannotated gels to large databases of images integrated with protein and nucleic acid sequence banks.

One example of an integrated proteome database is the suite of SWISS-PROT, SWISS-2DPAGE and SWISS-3DIMAGE databases (Appel *et al.*, 1993; Appel *et al.*, 1994; Appel, Bairoch and Hochstrasser, 1994; Bairoch and Boeckmann, 1994). The features of these three databases are listed in *Table 4*. SWISS-PROT, SWISS-2DPAGE and SWISS-3DIMAGE are accessible through the World Wide Web

**Table 4:** The SWISS-PROT, SWISS-2DPAGE and SWISS-3DIMAGE suite of crosslinked databases. All three databases are accessible through the World Wide Web, at URL address: <http://expasy.hugobuch/>

	SWISS-PROT	SWISS-2DPAGE	SWISS-3DIMAGE
Information	Text entries of sequence data. Citation information. taxonomic data, 38, 303 entries in Release 29	2-D gel images of: human liver, plasma, HepG2, HepG2 secreted proteins, red blood cell, lymphoma, cerebrospinal fluid, macrophage like cell line, erythroleukemia cell, platelet	Collection of 330 3-D images of proteins
Annotations	Protein function. Post translational modifications. Domains. Secondary structure. Quaternary structure. Diseases associated with protein. Sequence conflicts	Gel images where protein is found. How protein identified. Protein pI and MW. protein number. normal and pathological variants	All annotation is available in SWISS- PROT
Cross- Referenced Databases	SWISS-2DPAGE SWISS-3DIMAGE EMBL, PIR, PDB, OMIM, PROSITE, Medline, Flybase, GCRDB, MaizeDB, WonnPep, DictyDB	SWISS-PROT and all other databases accessible through SWISS-PROT	SWISS-PROT and all other databases accessible through SWISS-PROT
Other Features	Navigation to other SWISS databases achieved by selecting entries with computer mouse	Gel images show position of identified proteins, or region of gel where protein should appear	Mono and stereo images available. Images can be transferred to local computer image viewing programs

(Berners-Lee *et al.*, 1992), allowing any computer connected to the internet to access the stored information and images. Navigation within and between the three databases is seamless, as all potential crosslinks are highlighted as hypertext on the display and can be selected with a computer mouse. From these databases, detailed information about a protein, including amino acid sequence and known post-translational modifications, can be obtained, the precise protein spot it corresponds to on a reference gel image can be viewed if known, and the 3-D structure of the molecule can be seen if available. References to nucleic acid and other databases are also given to provide access to information stored elsewhere.

Organism databases, containing detailed protein and nucleic acid information about a species, are becoming common as genome and proteomic projects progress. These differ from nucleic acid or protein sequence databases like GenBank or SWISS-PROT because they are image based, and contain information about chromosomal map positions, transcription of genes, and protein expression patterns. The *Escherichia coli* gene-protein database (VanBogelen, Hutton and Neidhardt, 1990; VanBogelen and Neidhardt, 1991; VanBogelen *et al.*, 1992), known as the ECO2DBASE, is one example. It contains gene and protein names, 2-D gel spot information (including pI and MW estimates, and spot identification), genetic information (GenBank or EMBL codes, chromosomal location, location on Kohara clones (Kohara, Akiyama, and Isono, 1987), transcription direction of genes), and protein regulatory information (level of protein expression under different growth regimes, member of regulon or stimulon). All entries in the ECO2DBASE are also cross-referenced to the SWISS-PROT database (Bairoch and Boeckmann, 1994). It is anticipated that organism databases will soon become a standard means of storing all available information about a particular species. However there is currently no consistent manner in which organism databases are assembled, which may hamper comparisons in the future.

### Identification and characterisation of proteins from 2-D gels

The number of proteins identified on a 2-D reference map determines its usefulness as a research and reference tool. As most reference maps have only a small proportion of proteins identified, a major aim of current proteome projects is to screen many proteins from 2-D maps, in order to define them as 'known' in current nucleic acid and protein databases, or as 'unknown'. Protein identification assists in confirmation of DNA open reading frames, and provides focus for DNA sequencing projects and protein characterisation efforts by pointing to proteins that are novel. Since there may be 3000–4000 proteins from a single 2-D map that require identification, the challenge in protein screening is to identify proteins quickly, with a minimum of cost and effort.

Traditionally, proteins from 2-D gels have been identified by techniques such as immunoblotting, N-terminal microsequencing, internal peptide sequencing, comigration of unknown proteins with known proteins, or by overexpression of homologous genes of interest in the organism under study (Matsudaira, 1987; Rosenfeld *et al.*, 1992; VanBogelen *et al.*, 1992; Celis *et al.*, 1993; Honore *et al.*, 1993; Garrels *et al.*, 1994). Whilst these techniques are powerful identification tools, they are too expensive or time and labour intensive to use in mass screening programs. A hierarchical approach to mass protein identification has been recently suggested as an

alter  
use c  
mass  
slow  
the c  
of th  
macl  
cons  
tech  
ident

PROT

Ther  
identi  
This  
to id  
The  
radi  
*al.*  
chro  
1988  
1994  
phor  
radi

**Table 5:** Hierarchical analysis for mass screening of 2-D separated proteins blotted to membranes. Rapid and inexpensive techniques are used as a first step in protein identification, and slower, more expensive techniques are then used if necessary. Table modified from Wasinger *et al.*, 1995.

Order	Identification technique	References
1	Amino acid analysis	Jungblut <i>et al.</i> , 1992; Shaw, 1993; Hohmann, Houthaeve and Sander, 1992; Jungblut <i>et al.</i> , 1994; Wilkins <i>et al.</i> , 1995
2	Amino acid analysis with N-terminal sequence tag	Wilkins <i>et al.</i> , submitted
3	Peptide-mass fingerprinting	Henzel <i>et al.</i> , 1993; Pappin, Hourup and Biczsky, 1993; James <i>et al.</i> , 1993; Mann, Hourup and Roepstorff, 1993; Yates <i>et al.</i> , 1993; Mann <i>et al.</i> , 1994; Sutton <i>et al.</i> , 1995
4	Combination of amino acid analysis and peptide mass fingerprinting	Cordwell <i>et al.</i> , 1995; Wasinger <i>et al.</i> , 1995
5	Mass spectrometry sequence tag	Mann and Wilm, 1994
6	Extensive N-terminal Edman microsequencing	Matsudaira, 1987
7	Internal peptide Edman microsequencing	Rosenfeld <i>et al.</i> , 1992; Hellman <i>et al.</i> , 1995
8	Microsequencing by mass spectrometry (electrospray ionisation, post-source decay MALDI-TOF)	Johnson and Walsh, 1992
9	Ladder sequencing	Barile-Jones <i>et al.</i> , 1994

alternative to traditional approaches (Table 5; Wasinger *et al.*, 1995). This involves the use of rapid and cheap identification tools such as amino acid analysis and peptide mass fingerprinting as first steps in protein identification, followed by the use of slower, more expensive and time consuming identification procedures if necessary. In the construction of this hierarchy the analysis time, cost per sample and the complexity of the data created has been considered, as whilst some techniques require little machine time per sample, the analysis of data can be quite involved and time consuming. Amino acid analysis and peptide mass-fingerprinting based identification techniques in the hierarchy are discussed in detail below. For review of other protein identification techniques in Table 5, see Patterson (1994) and Mann (1995).

#### PROTEIN IDENTIFICATION BY AMINO ACID COMPOSITION

There has been a revival of interest in the use of amino acid composition for identification of proteins from 2-D gels after early work by Eckerskorn *et al.* (1988). This technique uses a protein's idiosyncratic amino acid composition profile in order to identify it by comparison with theoretical compositions of proteins in databases. The amino acid composition of proteins can be determined by differential metabolic radiolabelling and quantitative autoradiography after 2-D electrophoresis (Garrels *et al.*, 1994; Frey *et al.*, 1994), or by acid hydrolysis of membrane-blotted proteins and chromatographic analysis of the resulting amino acid mixture (Eckerskorn *et al.*, 1988; Touse *et al.*, 1989; Gharahdaghi *et al.*, 1992; Jungblut *et al.*, 1992; Wilkins *et al.*, 1995). As differential metabolic labelling experiments require X-ray film or phosphor-image plate exposures of up to 140 days, and can only be undertaken with easily radiolabelled samples, the technique is not as rapid or widely applicable as chromato-

## Spot: ECOLI-21M

\*\*\*\*\*

## Composition:

Asx: 13.2 Glx: 10.4 Ser: 5.7 His: 0.7  
 Gly: 5.4 Thr: 3.6 Ala: 6.7 Pro: 7.9  
 Tyr: 1.3 Arg: 5.0 Val: 8.0 Met: 0.3  
 Ile: 5.9 Leu: 8.0 Phe: 13.3 Lys: 4.4

pI estimate: 6.89 Range searched: (6.64, 7.14)  
 Mw estimate: 16200 Range searched: (13440, 20160)

Closest SWISS-PROT entries for the species ECOLI matched by AA composition:

Rank	Score	Protein	pI	Mw	Description
1	24	PYRI_ECOLI	6.84	16989	<b>ASPARTATE CARBAMOYLTRANSFERASE</b>
2	39	COAA_ECOLI	6.32	16359	PANTOTHENATE KINASE (EC 2.7.1.33)
3	40	META_ECOLI	5.06	35713	HOMOSERINE O-SUCCINYLTRANSFERASE
4	42	CADC_ECOLI	5.52	57812	TRANSCRIPTIONAL ACTIVATOR CADC.
5	43	HLVC_ECOLI	8.56	19769	HEMOLYSIN C. PLASMID.

Closest SWISS-PROT entries for ECOLI with pI and Mw values in specified range:

Rank	Score	Protein	pI	Mw	Description
1	24	PYRI_ECOLI	6.84	16989	<b>ASPARTATE CARBAMOYLTRANSFERASE</b>
2	102	TRAJ_ECOLI	6.73	17921	TRAJ PROTEIN.
3	112	YAJG_ECOLI	6.79	19028	HYPOTHETICAL LIPOPROTEIN YAJG.
4	140	YFJB_ECOLI	6.83	14945	HYPOTHETICAL 14.9 KD PROTEIN IN GRPE
5	142	YAHA_ECOLI	7.06	14726	HYPOTHETICAL PROTEIN IN BETT 3' REGION

Figure 4. Computer printout from ExPASy server where the empirical amino acid composition, estimated pI and MW of a protein from a 2-D reference map of *E. coli* were matched against all entries in SWISS-PROT for *E. coli*. The correct identification, aspartate carbamoyltransferase, is shown in bold. Low scores indicate a good match. Note how matching within a defined pI and MW range (lower set of proteins) has greatly increased the score difference between the first and second ranking proteins. This score difference gives high confidence in the identification, and is only observed where the top ranking protein is the correct identification (Wilkins *et al.*, 1995).

graphy-based analysis. Proteins blotted to PVDF membranes can be hydrolysed in 1 h at 155°C, amino acids extracted in a single brief step, and each sample automatically derivatised and separated by chromatography in under 40 minutes (Wilkins *et al.*, 1995; Ou *et al.*, 1995). In this manner, one operator can routinely analyse 100 proteins per week on one HPLC unit. This technology lends itself to automation, and it is anticipated that instruments with even greater sample throughput will be developed. When proteins have been prepared by micropreparative 2-D electrophoresis (Hanash *et al.*, 1991; Bjellqvist *et al.*, 1993b), blotted to a PVDF membrane and stained with amido black, any visible protein spot is of sufficient quantity for amino acid analysis (Cordwell *et al.*, 1995; Wasinger *et al.*, 1995; Wilkins *et al.*, 1995).

After the amino acid composition of a protein has been determined, computer programs are used to match it against the calculated compositions of proteins in databases (Eckerskorn *et al.*, 1988; Sibbald, Sommerfeldt and Argos, 1991; Jungblut *et al.*, 1992; Shaw, 1993; Hobohm, Houthaeve and Sander, 1994; Wilkins *et al.*, 1995). Matching is usually done with only 15 or 16 amino acids, as cysteine and

Figure:  
same as  
acid com  
PROT  
for this  
large as  
the cur  
protein

trypto  
to the  
The co  
a scor  
restric  
1994;  
*et al.*  
match  
in Fig  
refer  
romy  
lympl  
*et al.*

PROT:  
SEQU  
When



Spot: ECOLI-ACC

\*\*\*\*\*

## Composition:

Asx: 5.4 Glx: 10.8 Ser: 4.1 His: 2.7  
 Gly: 12.2 Thr: 3.8 Ala: 11.9 Pro: 3.2  
 Tyr: 6.0 Arg: 3.7 Val: 9.5 Met: 0.6  
 Ile: 5.0 Leu: 8.2 Phe: 3.2 Lys: 4.9

pI estimate: 5.99 Range searched: ( 5.74, 6.24)

Mw estimate: 45000 Range searched: (36000, 54000)

Closest SWISS-PROT entries for ECOLI with pI and Mw values in specified range:

Rank	Score	Protein	pI	Mw	N-terminal Seq.
1	21	GLYA_ECOLI	6.03	45316	M L K R E
2	32	YJGB_ECOLI	5.86	36502	M S H I K
3	38	GART_ECOLI	5.78	45774	M S N S K
4	44	YIHS_ECOLI	5.86	48018	M R I K Y
5	45	DHE4_ECOLI	5.98	48581	M D Q T Y
6	46	ARGO_ECOLI	5.79	43765	M A I E Q
7	46	NJPB_ECOLI	5.78	37851	M N H S L
8	47	GLAC_ECOLI	5.98	49162	M L N R A
9	47	ACKA_ECOLI	5.85	43290	M S S K L
10	50	YJTH_ECOLI	6.01	37064	M E S R :

Figure 5. A PVDF protein spot from an *E. coli* 2-D reference map was sequenced for 4 cycles, and the same sample then subject to amino acid analysis. The N-terminal sequence was M L K R. When the amino acid composition of the spot, as well as estimated pI and MW, were matched against all entries in SWISS-PROT for *E. coli*, the above list of best matches was produced. N-terminal sequences are from SWISS-PROT for those entries. The top ranking identification of serine hydroxymethyltransferase (bold) did not show a large score difference between the first and second ranking proteins, giving little confidence in this being the correct protein identification. However, the sequence tag (M L K R) confirmed the identity of the protein as serine hydroxymethyltransferase.

tryptophan are destroyed during hydrolysis, asparagine and glutamine are deamidated to their corresponding acids, and proline is not quantitated in some analysis systems. The computer programs produce a list of best matching proteins, which are ranked by a score that indicates the match quality. Some programs allow matching to be restricted to specific 'windows' of MW and pI (Hobohm, Houthaeve and Sander, 1994; Wilkins *et al.*, 1995), and to protein database entries for one species (Jungblut *et al.*, 1992; Wilkins *et al.*, 1995). The use of such restrictions increases the power of matching. An example of protein identification by amino acid composition is shown in Figure 4. To date, amino acid composition has been used to identify proteins from reference maps of *Spiroplasma melliferum*, *Mycoplasma genitalium*, *E. coli*, *Saccharomyces cerevisiae*, *Dictyostelium discoideum*, human sera, human heart, human lymphocyte, and mouse brain (Cordwell *et al.*, 1995; Wasinger *et al.*, 1995; Wilkins *et al.*, 1995; Jungblut *et al.*, 1992, 1994; Garrels *et al.*, 1994; Frey *et al.*, 1994).

#### PROTEIN IDENTIFICATION BY AMINO ACID COMPOSITION AND N-TERMINAL SEQUENCE TAG

When samples from 2-D gels are not unambiguously identified by amino acid

composition, pI and MW, often the correct identification of that protein is amongst the top rankings of the list (Hohohm, Houthaeve and Sander, 1994; Cordwell *et al.*, 1995; Wilkins *et al.*, 1995). Taking advantage of this observation, we have used the mass spectrometry 'sequence tag' concept (Mann and Willen, 1994) in developing a combined Edman degradation and amino acid analysis approach to protein identification (Wilkins *et al.*, submitted). This involves the N-terminal sequencing of PVDF-blotted proteins by Edman degradation for 3 or 4 cycles to create a 'sequence tag', following which the same sample is used for amino acid analysis. As only a few amino acids are removed from the protein, its composition is not significantly altered. Furthermore, since only a small amount of protein sequence is required, fast but low repetitive yield Edman degradation cycles can be used. Modifications to current procedures should allow 3 cycles to be completed in 1 h, thereby allowing the screening of 100 or more proteins per week on one automated, multi-cartridge sequencer. Amino acid composition, pI and MW of proteins are matched against databases as described above, and N-terminal sequences of best matching proteins are checked with the 'sequence tag' to confirm the protein identity (Figure 5). This technique will be less useful when proteins are N-terminally blocked, but as only a few N-terminal amino acids are susceptible to the acetyl, formyl, or pyroglutamyl modifications that cause blockage, this may itself provide useful information for sequence tag identification. A strength of N-terminal sequence tag and amino acid composition protein identification is that data generated are quickly and easily interpreted.

#### PROTEIN IDENTIFICATION BY PEPTIDE MASS FINGERPRINTING

Techniques for the identification of proteins by peptide mass fingerprinting have recently been described (Henzel *et al.*, 1993; Pappin, Hojrup and Bleasby, 1993; James *et al.*, 1993; Mann, Hojrup and Roepstorff, 1993; Yates *et al.*, 1993; Mortz *et al.*, 1994; Sutton *et al.*, 1995). This involves the generation of peptides from proteins using residue-specific enzymes, the determination of peptide masses, and the matching of these masses against theoretical peptide libraries generated from protein sequence databases. As proteins have different amino acid sequences, their peptides should produce characteristic 'fingerprints'.

The first step of peptide mass fingerprinting is protein digestion. Proteins within the gel matrix or bound to PVDF can be enzymatically digested *in situ*, although *in situ* gel digests are reported to produce more enzyme autodigestion products, which complicate subsequent peptide mass analysis (James *et al.*, 1993; Rasmussen *et al.*, 1994; Mortz *et al.*, 1994). The enzyme of choice for digestion is currently trypsin (of modified sequencing grade), but other enzymes (Lys-C or *S. aureus* V8 protease) have also been used (Pappin, Hojrup and Bleasby, 1993). To maximise the number of peptides obtained, it is desirable for protein samples to be reduced and alkylated prior to digestion (Mortz *et al.*, 1994; Henzel *et al.*, 1993). This ensures that all disulfide bonds of the protein are broken, and produces protein conformations that are more amenable to digestion. Surprisingly, chemical digestion methods such as cyanogen bromide (methionine specific), formic acid (aspartic acid specific), and 2-(2'-nitrophenylsulfonyl)-3-methyl-3-bromoindolenine (tryptophan specific) have not been explored as means of peptide production for mass fingerprinting, even though they are rapid and may circumvent some problems associated with enzyme digestions.

(Nikodem and Fresco, 1979; Crimmins *et al.*, 1990; Vanfleteren *et al.*, 1992).

After proteins are digested, peptide masses are determined by mass spectrometry. Direct analysis of peptide mixtures can be achieved by electrospray ionisation mass spectrometry, plasma desorption mass spectrometry, or matrix assisted laser desorption ionization (MALDI) mass spectrometry techniques. MALDI is preferable because of its higher sensitivity and greater tolerance to contaminating substances from 2-D gels (James *et al.*, 1993; Moritz *et al.*, 1994; Pappin, Hojrup and Bleasby, 1993). Furthermore, recent modifications to sample preparation methods have largely solved early difficulties experienced with the calibration of MALDI spectra (Moritz *et al.*, 1994; Vorm and Mann, 1992; Vorm, Roepstorff and Mann, 1994). The high sensitivity of mass spectrometry allows a small fraction of a digest of a 1 µg protein spot to be used for analysis, and analysis itself is complete in a few minutes.

A major challenge associated with peptide mass fingerprinting is data interpretation prior to computer matching against libraries of theoretical peptide digests. Spectra must be examined carefully to determine which peaks represent peptide masses of interest, as there are often enzyme autodigestion products and contaminating substances present (Henzel *et al.*, 1993; Moritz *et al.*, 1994; Rasmussen *et al.*, 1994). Furthermore, if protein alkylation and reduction has not been undertaken prior to protein digestion, peptide sequence coverage may be poor (40% to 70%), with some masses present representing disulfide bonded peptides originally present in the protein (Moritz *et al.*, 1994). For eukaryotes, a serious issue is the alteration of peptide masses by the presence of post-translational modifications (Table 6). The mass of the unmodified peptide alone can be very difficult to determine. Two artifactual modifications introduced by electrophoresis, an acrylamide adduct to cysteine and the oxidation of methionine, are also known to alter peptide masses (Le Maire *et al.*, 1993; Hess *et al.*, 1993).

Table 6: Masses of some common post-translational modifications. Peptides carrying post-translational modifications complicate data analysis for peptide mass fingerprinting protein identification. This is especially so for protein glycosylation, which involves many different combinations of the hexosamines, hexoses, deoxyhexoses, and sialic acid

Post-translational modification	Mass change
Acetylation	- 42.04
* Acrylamide adduct to cysteine	- 71.00
Carboxylation of Asp or Glu	- 71.00
Deamidation of Asn or Gln	- 24.01
Disulfide bond formation	- 0.98
Deoxyhexoses (Fuc)	- 2.02
Formylation	146.14
Hexosamines (GlcN, GalN)	- 28.01
Hexoses (Glc, Gal, Man)	- 161.16
Hydroxylation	- 162.14
N-acetylhexosamines (GlcNAc, GalNAc)	- 16.00
* Oxidation of Met	- 203.19
Phosphorylation	- 16.00
Pyroglutamic acid formed from Gln	- 74.98
Sialic acid (NeuNAc)	- 17.03
Sulfation	- 291.26
	- 80.06

Table modified from Finnigan LASERBIAT application data sheet 5.

Asterisk \* shows modifications that can arise artifactually from the 2-D electrophoresis process

A number of computer programs are available for matching peptide masses against databases (reviewed in Cottrell, 1994). Matching is usually undertaken in an interactive manner, whereby peaks of mass 500–3000 Da are selected and matched under various search parameters including MW of protein, mass accuracy of peptides, and number of missed enzyme cleavages allowed (Henzel *et al.*, 1993; Moritz *et al.*, 1994; Rasmussen *et al.*, 1994). The correct protein identity is the protein which has the most peptide masses in common with the unknown sample. Identities have been established with as few as three peptides, but unambiguous identification is thought to require a mass spectrometric map covering most peptides of the protein (Moritz *et al.*, 1994; Yates *et al.*, 1993). To date, peptide mass fingerprinting of proteins has been undertaken from the human myocardial protein and keratinocyte maps, from an *E. coli* 2-D gel, and from reference maps of *Spiroplasma melliferum* and *Mycoplasma genitalium* (Sutton *et al.*, 1995; Rasmussen *et al.*, 1994; Henzel *et al.*, 1993; Cordwell *et al.*, 1995; Wasinger *et al.*, 1995), although the technique is most powerful when used in combination with another protein identification technique (Rasmussen *et al.*, 1994; Cordwell *et al.*, 1995).

#### MASS SPECTROMETRY SEQUENCE TAGGING

An extension of peptide mass fingerprinting has recently been described, called peptide sequence tagging (Mann and Wilm, 1994; Mann, 1995). This uses tandem mass spectrometry (MS/MS) to initially determine the mass of peptides, then subject them to fragmentation by collision with a gas, and finally determine the mass of fragments. The resulting spectra gives information about a peptide's amino acid sequence. The fragmentation masses of peptides can rarely be used to assign a complete sequence, but it usually allows a short 'sequence tag' of 2 or 3 amino acids to be determined. This sequence tag and the original peptide mass is matched by computer against a database, providing a likely identity of the peptide and the protein it came from. The major drawback for this technique as a mass screening tool is the complexity of the mass data generated and the high level of expertise required for its interpretation. Nevertheless, it represents a useful new protein identification method which greatly increases the power of peptide mass fingerprinting protein identification.

#### Cross-species protein identification

Protein sequence databases continue to grow at a rapid rate, yet it is not widely appreciated that close to 90% of all information contained in current protein databases comes from only 10 species (A. Bairoch, Pers. Comm.). Fortunately, this information can be used to study proteomes of organisms that are poorly defined at the molecular level, via 2-D electrophoresis and 'cross-species' protein identification (Cordwell *et al.*, 1995; Wasinger *et al.*, 1995). This approach allows proteins from reference maps of many different species to be identified without the need for the corresponding genes to be cloned and sequenced. This is particularly true for 'housekeeping' proteins, such as enzymes involved in glycolysis, DNA manipulation and protein manufacture, which are highly conserved across species boundaries. Proteins that cannot be identified across species boundaries can then become the focus of further protein characterisation and DNA sequencing efforts.

Figure  
and try  
be con  
identit  
and H  
all en  
protein  
progra  
match  
apudip

A)

Protein: APA1\_HUMAN

\*\*\*\*\*

Asx: 8.4 Cix: 19.3 Ser: 6.3 His: 1.3  
 Gly: 4.2 Tex: 4.3 Ala: 8.0 Pro: 4.2  
 Tyr: 2.9 Lrg: 6.7 Val: 5.5 Met: 1.3  
 Cile: 0.0 Ieu: 15.5 Phe: 2.5 Lys: 8.8

pI Range: no range specified

Mw Range: no range specified

The closest SWISS-PROT entries are:

Rank	Score	Protein	(pI	Mw)	Description
1	0	APA1_HUMAN	5.27	28078	APOLIPOPROTEIN A-I.
2	4	APA1_MACFA	5.43	28005	APOLIPOPROTEIN A-I.
3	10	APA1_RABIT	5.15	27836	APOLIPOPROTEIN A-I.
4	14	APA1_BOVIN	5.36	27549	APOLIPOPROTEIN A-I.
5	14	APA1_CANFA	5.10	27467	APOLIPOPROTEIN A-I.
6	18	APA1_MOUSE	5.42	27922	APOLIPOPROTEIN A-I.
7	26	APA1_PIG	5.19	27598	APOLIPOPROTEIN A-I.
8	27	APA1_CHEEK	5.26	27966	APOLIPOPROTEIN A-I.
9	37	DYTA_CHEEK	5.44	217742	DYNASTIN, 117 KD ISOFORM.
10	29	APA1_HUMAN	5.18	43374	APOLIPOPROTEIN A-IV.

B)

Reagent: Trypsin MW filter: 10k

Scan using fragment mws of:

1953 1933 1731 1613 1401 1387  
 1301 1283 1252 1235 1231 1215  
 1031 896 873 831 813 781  
 732 704

No. of database entries scanned = 72018

1	APA1_HUMAN	APOLIPOPROTEIN A-I (APO-AI). - HOMO SAPIENS
2	APA1_MACFA	APOLIPOPROTEIN A-I (APO-AI). - MACACA FASCICULARIS
3	APA1_PAPHA	APOLIPOPROTEIN A-I (APO-AI). - PAPIO HAMADRYAS
4	B41845	csf B - Treponema denticola
5	APA1_CANFA	APOLIPOPROTEIN A-I (APO-AI). - CANIS FAMILIARIS (DOG).
6	S30947	hypothetical protein 1 - Azotobacter vinelandii
7	H520_PEA	CHLOROPLAST HEAT SHOCK PROTEIN PRECURSOR. - PISUM SATIVUM
8	S20724	Tropomyosin - African clawed frog
9	HIVV1354	HIVV1354 premature term. at 793 - Human immunodeficiency
10	TRAI_ECOLI	TRAJ PROTEIN. - ESCHERICHIA COLI.

Figure 6. Theoretical cross-species matching of human apolipoprotein A-I by amino acid composition and tryptic peptides. When an unknown protein is analysed, best ranking proteins from both techniques can be compared. If the same protein type is observed in both lists, there is high confidence in this being the identity of the unknown molecule (Cordwell *et al.*, 1995). (A) Output of ExPASy server (Appel, Bairoch and Hochstrasser, 1993) where the true amino acid composition of apolipoprotein A-I was matched against all entries in the SWISS-PROT database, without pI or MW windows. Seven of the top 10 matching proteins were apolipoprotein A-I of different species. (B) Output of MOWSE peptide mass fingerprinting program (Pappin, Hojrup and Bleasby, 1993) where true tryptic peptides of human apolipoprotein A-I were matched against the OWL database, using MW window of 10%. Four of the top ten matching proteins were apolipoprotein A-I from different species.

Rapid cross-species identification of proteins from 2-D reference maps can be undertaken with amino acid composition or peptide mass fingerprinting methods (Figure 6), but these techniques alone may not identify proteins unambiguously when phylogenetic cross-species distances are great or analysis data is of poor quality (Yates *et al.*, 1993; Shaw, 1993; Cordwell *et al.*, 1995). However, very high confidence in protein identities can be achieved when lists of best-matching proteins generated by both techniques are compared (Cordwell *et al.*, 1995; Wasinger *et al.*, 1995). The correct identification is found when the same protein is ranked highly in lists of best matches generated by both techniques. This method has allowed approximately 120 proteins from the reference map of the mollicute *Spiroplasma melliferum*, representing approximately one quarter of the proteome, to be confidently identified by reference to protein information from other species (S. Cordwell, Personal Communication). When cross-species protein identification is to be undertaken, it should be noted that the molecular weight of a protein type across species is usually highly conserved, but that protein pI can vary by more than 2 units (Cordwell *et al.*, 1995). Accurate molecular weight determination by direct mass spectrometry of proteins blotted to PVDF (Eckerskorn *et al.*, 1992) should therefore be a useful additional parameter for cross-species protein identification.

#### CHARACTERISATION OF POST-TRANSLATIONAL MODIFICATIONS

Many proteins are modified after translation. Such post-translational modifications, including glycosylation, phosphorylation, and sulfation (see Table 6), are usually necessary for protein function or stability. Some abnormal modifications are associated with disease (Duthel and Revol, 1993; Ghosh *et al.*, 1993; Yamashita *et al.*, 1993). In proteome studies, post-translational modifications can be examined on all proteins present, or on individual spots. Studies on all proteins provide an indication of which proteins may carry a certain type of modification. For example, 2-D gel analysis of cell cultures grown in the presence of [<sup>3</sup>H] mannose or [<sup>32</sup>P] phosphate gives an indication of which proteins carry glycans containing mannose, and which proteins are phosphorylated (Garrels and Franza, 1989). Lectin binding studies of 2-D gels blotted to PVDF or nitrocellulose provide information on the saccharides, if any, that are carried by proteins present (Gravel *et al.*, 1994).

When individual proteins of interest carrying post-translational modifications have been found, micropreparative 2-D electrophoresis can be used to purify them in microgram quantities (Hanash *et al.*, 1991; Bjellqvist *et al.*, 1993b). If protein isoforms of similar MW and pI are to be studied, focusing with narrow range pI gradients (1 pH unit) can provide greater separation and resolution. After electrophoresis, the type and degree of protein phosphorylation can be investigated (Murthy and Iqbal, 1991; Gold *et al.*, 1994), monosaccharide composition can be determined (Weitzhandler *et al.*, 1993; Packer *et al.*, 1995), and the structure and exact site of glycoamino acids can be investigated by either Edman degradation based techniques or by mass spectrometry (Pisano *et al.*, 1993; Huberty *et al.*, 1993; Carr, Huddleston and Bean, 1993). With further development of rapid techniques, investigation of phosphorylation and monosaccharides by chromatographic or mass spectrometric means is likely to become a routine step in the characterisation of post-translational modifications of proteins from reference maps.

The  
M  
com  
this  
plasm  
Wasi  
maps  
speci  
and si

Table  
PROT  
referen  
from B  
1996

Species

*Mycop.  
Escheri  
Sacchar  
Dietrich  
Arabid.  
Caenor  
Himno*

The  
under  
becau  
hundr  
estim  
to tiss  
protei  
electr  
ism c  
accel  
are ur  
post-t  
differ  
useful

## The status of proteome projects

Many technical aspects of proteome research have already been discussed in this review, but an overview of the status of proteome projects has not yet been presented. Advances in proteome projects will initially rely on progress in genome sequencing initiatives, to enable an identity, amino acid sequence, or function to be assigned to each protein spot. Table 7 shows genome size, proteome size, and the number of proteins already defined for a number of model organisms. This indicates that whilst genome sequencing programs for *E. coli* and *S. cerevisiae* are advanced, the massive size of some other genomes (and especially the human genome) means that their complete nucleotide sequences are unlikely to be available for many years. Because of this, 2-D reference maps and proteome projects of single cell organisms like *Mycoplasma* sp., *E. coli* and *S. cerevisiae* will be the most detailed (Cordwell *et al.*, 1995; Wasinger *et al.*, 1995; Vanbogelen *et al.*, 1992; Garrels *et al.*, 1994), and complete maps of other organisms will take longer to construct. However, the use of cross-species protein identification techniques will allow proteomes of many prokaryotes and simple eukaryotes to be partially defined in reference to *E. coli* and *S. cerevisiae*.

Table 7: Estimated genome size, estimated proteome size, number of protein sequences in SWISS-PROT Release 31 (March, 1995), and approximate number of proteins of known identity on 2-D reference maps for some model organisms. Genome size data from Smith (1994), and total protein data from Bird (1995). Genome sequencing projects of *E. coli* and *S. cerevisiae* will probably be complete in 1996.

Species Name	Haploid genome size (million bp)	Estimated proteome size (total proteins)	Protein entries in SWISS-PROT	Proteins annotated on 2-D Maps
<i>Mycoplasma</i> species	0.6-0.8	400-600	100	> 100
<i>Escherichia coli</i>	4.8	4000	3170	> 300
<i>Saccharomyces cerevisiae</i>	12.5	6000	3160	> 100
<i>Drosophila melanogaster</i>	70	12500	202	-
<i>Arabidopsis thaliana</i>	70	14000	270	-
<i>Caenorhabditis elegans</i>	80	17000	703	-
<i>Homo sapiens</i>	2900	60000-80000	3326	> 1000

The study of vertebrate proteomes and vertebrate development is a phenomenal undertaking in comparison to the investigation of single cell organisms. This is because vast numbers of proteins are developmentally expressed, each body tissue has hundreds of unique proteins, and there are numerous tissue types. However, it is estimated that at least 35% of proteins in vertebrate cells will be conserved from tissue to tissue, constituting the 'housekeeping' proteins (Bird, 1995), with the remainder of proteins constituting a set that are specific to a cell type. Providing that standardised electrophoretic conditions are used, reference maps from many tissues of one organism can be superimposed in gel databases (e.g. Hochstrasser *et al.*, 1992). This accelerates the definition of the 'housekeeping' proteins, as well as sets of proteins that are unique to different tissue types. Such studies may, however, be complicated by post-translational modifications, which can differ on the same gene product in different tissues. Proteins that remain unknown after identification procedures will be useful in providing focus for nucleic acid sequencing initiatives.

## FUTURE DIRECTIONS OF PROTEOME PROJECTS

This review has described recent advances in the area of proteome research. It has illustrated how new developments of older techniques (2-Delectrophoresis and amino acid analysis) as well as the applications of new technology (mass spectrometry) have greatly widened the choice of tools the biologist and protein chemist has for the separation, identification and analysis of complex mixtures of proteins. This has made possible the establishment of detailed reference maps for organisms, which are becoming the method of choice for the definition of tissues or whole cells, and the investigation of gene expression therein.

Proteome projects are already impacting on the dogma of molecular biology that DNA sequence constitutes the definition of an organism. For example, the proteomes of different tissues of a single organism are often significantly different. Similarly, cross-species identification of proteins (for example the identification of proteins from *Candida albicans* by comparison with *S. cerevisiae*) can open up studies on organisms that are poorly molecularly defined. As cross-species identification can proceed at a pace orders of magnitude faster than a genome project in terms of defining the gene and protein complement of organisms, the need for the DNA sequencing of genomes will be avoided, and emphasis placed on those found to be novel.

Just as genome sequencing is not an end in itself, neither is an annotated 2-D protein reference map of an organism, nor indeed the identification of proteins in a proteome. So whilst an immediate aim of proteome projects is to screen proteins in reference maps, this will lead to expression studies and characterisation of post-translational modifications. The challenge that then needs to be addressed is the investigation of structure and function of proteins in a proteome. The magnitude of this is illustrated by the fact that over half the open reading frames identified in *S. cerevisiae* chromosome III were initially of no known function (Oliver *et al.*, 1992). Structural and functional studies will be an undertaking just as formidable as genome studies are now and proteome projects are becoming, but will lead to an unimaginably detailed understanding of how living organisms are constructed and how they operate.

## Acknowledgements

MRW is the recipient of an Australian Postgraduate Research Award. AAG, MRW, IHS and KIW acknowledge assistance for proteome projects through Macquarie University Research Grants, the Australian Research Council, the Australian National Health and Medical Research Council, Beckman Instruments and GBC Scientific Equipment. DH acknowledges the financial support of a Montux Foundation Grant and the Swiss National Fund for Scientific Research (Grant # 31-33658.92). We thank colleagues who supplied work that was 'In Press' during the writing of this review.

## References

- ANDERSON, N.L., HOFMANN, J.P., GEMMELL, A. AND TAYLOR, J. (1984). Global approaches to quantitative analysis of gene-expression patterns observed by use of two-dimensional gel electrophoresis. *Clinical Chemistry*, 30, 2031-2036.



- ANDERSON, N.L. AND ANDERSON, N.G. (1991). A 140-dimensional gel database of human plasma proteins. *Electrophoresis*, 12, 852-906.
- ANDERSON, N.L., ESOLIER-BLASCO, R., HOFMANN, J.P. AND ANDERSON, N.G. (1991). A two-dimensional gel database of rat liver proteins useful in gene regulation and drug effects studies. *Electrophoresis*, 12, 907-930.
- ANDERSON, N.L., COPPLE, D.C., BENDELE, R.A., PROBST, G.S., RICHARDSON, F.C. (1992). Covalent protein modifications and gene expression changes in rodent liver following administration of methapyrilene: a study using two-dimensional electrophoresis. *Fundamental and Applied Toxicology*, 18, 570-580.
- APPEL, R.D., BAIROCH, A. AND HOCHSTRASSER, D.F. (1994). A new generation of information retrieval tools for biologists: the example of the ExPASy WWW server. *Trends in Biochemical Sciences*, 19, 256-260.
- APPEL, R.D., HOCHSTRASSER, D.F., FUNK, M., VARGAS, J.R., PELIGRINI, C., MÜLLER, A.F. AND SCHERRER, J.R. (1991). The MELANIE project: from a biopsy to automatic protein map interpretation by computer. *Electrophoresis*, 12, 722-735.
- APPEL, R.D., SANCHEZ, J.-C., BAIROCH, A., GOLAZ, O., MIL, M., VARGAS, J.R. AND HOCHSTRASSER, D.F. (1993). SWISS-2DPAGE: a database of two-dimensional gel electrophoresis images. *Electrophoresis*, 14, 1322-1326.
- APPEL, R.D., SANCHEZ, J.-C., BAIROCH, A., GOLAZ, O., RAVIER, F., PASQUALI, C., HUGHES, G. AND HOCHSTRASSER, D.F. (1994). The SWISS-2DPAGE database of two-dimensional polyacrylamide gel electrophoresis. *Nucleic Acids Research*, 22, 3581-3582.
- BAIROCH, A. AND BOECKMANN, B. (1994). The SWISS-PROT protein sequence databank: current status. *Nucleic Acids Research*, 22, 3578-3580.
- BAKER, C.S., CORBETT, J.M., MAY, A.J., YACOB, M.H. AND DUNN, M.J. (1992). A human myocardial two-dimensional electrophoresis database: protein characterisation by microsequencing and immunoblotting. *Electrophoresis*, 13, 723-726.
- BARTLET-JONES, M., JEFFERY, W.A., HANSEN, H.F. AND PAPPIN, D.J.C. (1994). Peptide ladder sequencing by mass spectrometry using a novel, volatile degradation reagent. *Rapid Communications in Mass Spectrometry*, 8, 737-742.
- BAUER, D., MÜLLER, H., REICH, J., RIEDEL, H., AHRENKIEL, V., WARTHOE, P. AND STRAUSS, M. (1993). Identification of differentially expressed mRNA species by an improved display technique (DDRT-PCR). *Nucleic Acids Research*, 21, 4272-4280.
- BERNERS-LEE, T.J., CAILLIAU, R., GROFF, J.F. AND POLLERMAN, B. (1992). *Electronic Networking: Research, Applications, and Policy*, 2, 52-58.
- BIRD, A.P. (1995). Gene number, noise reduction and biological complexity. *Trends in Genetics*, 11, 92-100.
- BJELLOVIST, B., EK, K., RIGHETTI, P.G., GIANAZZA, E., GORG, A., WESTERMEIER, R. AND POSTEL, W. (1982). Isoelectric focusing in immobilized pH gradients: principle, methodology and some applications. *Journal of Biochemical and Biophysical Methods*, 6, 317-339.
- BJELLOVIST, B., PASQUALI, C., RAVIER, F., SANCHEZ, J.-C. AND HOCHSTRASSER, D.F. (1993a). A nonlinear wide-range immobilized pH gradient for two-dimensional electrophoresis and its definition in a relevant pH scale. *Electrophoresis*, 14, 1357-1365.
- BJELLOVIST, B., SANCHEZ, J.-C., PASQUALI, C., RAVIER, F., PAQUET, N., FRITIGER, S., HUGHES, G.J. AND HOCHSTRASSER, D.F. (1993b). Micropreparative 2-D electrophoresis allowing the separation of milligram amounts of proteins. *Electrophoresis*, 14, 1375-1378.
- BJELLOVIST, B., HUGHES, G., PASQUALI, C., PAQUET, N., RAVIER, F., SANCHEZ, J.-C., FRITIGER, S. AND HOCHSTRASSER, D. (1993c). The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. *Electrophoresis*, 14, 1023-1031.
- BONNER, W.M. AND LASKEY, R.A. (1974). A film detection method for tritium-labeled proteins and nucleic acids in polyacrylamide gels. *European Journal of Biochemistry*, 46, 83-88.
- BOITTELL, T., GARRELS, J.L., FRANZA, B.R., MONARDO, P.J. AND LATTER, G.I. (1994). REF52 on global gel navigator: an internet-accessible two-dimensional gel electrophoresis database. *Electrophoresis*, 15, 1487-1490.
- BREWER, J., GRUND, E., HAGERLID, P., OLSSON, I. AND LIZANA, J. (1986). In *Electrophoresis '86* (M.J. Dunn, Ed.), pp. 226-229. VCH, Weinheim.

- FLY  
FRID  
GARE  
GARF  
GARK  
GARS  
GARY  
GELI  
GHAR  
GHOS  
GIOM  
GIOM  
GOLA  
GOLD  
GOLD  
GOOL  
GORC  
GORC

- Mass spectrometric analysis of blotted proteins after gel electrophoretic separation by matrix-assisted laser desorption/ionization. *Electrophoresis*, 13, 664-665.
- ECKERSKORN, C. AND LOTTSCHEICH, F. (1993). Structural characterization of blotting membranes and the influence of membrane parameters for electroblotting and subsequent amino acid sequence analysis of proteins. *Electrophoresis*, 14, 831-838.
- ER, K., BJELLOVIST, B.J. AND RIGHETTI, P.G. (1983). Preparative isoelectric focusing in immobilized pH gradients. I. General principles and methodology. *Journal of Biochemical and Biophysical Methods*, 8, 135-155.
- FEY, S.J., CARLSEN, J., MOSE LARSEN, P., JENSEN, U.A., KJELDSEN, K. AND HALVAND, S. (1994). Two-dimensional gel electrophoresis as a tool for molecular cardiology. Proceedings of the International Society for Heart Research 'XV' European Section Meeting, pp 9-16.
- FREY, J.R., KUHN, L., KETTMAN, J.R. AND LEFKOVITS, I. (1994). The amino acid composition of 350 lymphocyte proteins. *Molecular Immunology*, 31, 1219-1231.
- GARRELS, J.I. (1989). The QLEST system for quantitative analysis of two-dimensional gels. *Journal of Biological Chemistry*, 264, 5269-5282.
- GARRELS, J.I. AND FRANZA, B.R. (1989). The REF52 protein database. *Journal of Biological Chemistry*, 264, 5283-5295.
- GARRELS, J.I., FRANZA, B.R., CHANG, C. AND LATTER, G. (1990). Quantitative exploration of the REF52 protein database: cluster analysis reveals the major protein expression profiles in response to growth regulation, serum stimulation, and viral transformation. *Electrophoresis*, 11, 1111-1130.
- GARRELS, J.I., FUTCHER, B., KOBAYASHI, R., LATTER, I., SCHWENDER, B., VOLPE, T., WARNER, J.R. AND MCLAUGHLIN, C.S. (1994). Protein identification for a *Saccharomyces cerevisiae* protein database. *Electrophoresis*, 15, 1466-1486.
- GELFI, C., BOSSI, M.L., BJELLOVIST, B. AND RIGHETTI, P.G. (1987). Isoelectric focusing in immobilized pH gradients in the pH 10-11 range. *Journal of Biochemical and Biophysical Research Methods*, 15, 41-48.
- GHARAHDAAGHI, F., ATHERTON, D., DEMOTT, M. AND MISCH, S.M. (1992). Amino acid analysis of PVDF-bound proteins. In *Techniques in Protein Chemistry III* (R.H. Abele, Ed.), pp 249-260. Academic Press, San Diego.
- GHOSH, P., OKOH, C., LIL, Q.H. AND LAKSHMAN, M.R. (1993). Effects of chronic ethanol on enzymes regulating sialylation and desialylation of transferrin in rats. *Alcoholism: Clinical and Experimental Research*, 17, 576-579.
- GIOMETTI, C.S., GEMMELL, M.A., TOLLAKSEN, S.L. AND TAYLOR, J. (1991). Quantitation of human leukocyte proteins after silver staining: a study with two-dimensional electrophoresis. *Electrophoresis*, 12, 536-543.
- GIOMETTI, C.S., TAYLOR, J. AND TOLLAKSEN, S.L. (1992). Mouse liver protein database: a catalog of proteins detected by two-dimensional gel electrophoresis. *Electrophoresis*, 13, 970-991.
- GOLAZ, O., HUGHES, G.J., FRITIGER, S., PAQUET, N., BAIRICH, A., PASQUALI, C., SANCHEZ, J.C., TISSOT, J.D., APPEL, R.D., WALZER, C., BALANT, L. AND HOCHSTRASSER, D.F. (1993). Plasma and red blood cell protein maps: update 1993. *Electrophoresis*, 14, 1223-1231.
- GOLD, M.R., YUNGWIRTH, T., SUTHERLAND, C.L., INGHAM, R.J., VIANZON, D., CHIL, R., VAN-OOSTVEEN, I., MORRISON, H.D. AND AEBERSOLD, R. (1994). Purification and identification of tyrosine-phosphorylated proteins from lymphocytes stimulated through the antigen receptor. *Electrophoresis*, 15, 441-453.
- GOLDBERG, H.A., DOMENICUCCI, C., PRINGLE, G.A. AND SODEK, J. (1988). Mineral-binding proteoglycans of fetal porcine calvarial bone. *Journal of Biological Chemistry*, 263, 12092-12101.
- GOOLEY, A.A., MARSHCHALEK, R. AND WILLIAMS, K.L. (1992). Size polymorphisms due to changes in the number of O-glycosylated tandem repeats in the *Dictyostelium discoideum* glycoprotein PsA. *Genetics*, 130, 749-756.
- GORG, A., POSTEL, W. AND GUNTHER, S. (1988). The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis*, 9, 531-546.
- GORG, A., POSTEL, W., GUNTHER, S., WESER, J., STRAHLER, J.R., HANASHI, S.M., SOMERLOT, L.

- AND KUICK, R. (1985). Approach to stationary two-dimensional pattern: influence of focusing time and immobilized carrier ampholyte concentrations. *Electrophoresis*, 9, 37-46.
- GRAVEL, P., GOLAZ, O., WALZER, C., HOCHSTRASSER, D.F., TURLEZ, H., AND BALANT, L.P. (1994). Analysis of glycoproteins separated by two-dimensional gel electrophoresis using lectin blotting revealed by chemiluminescence. *Analytical Biochemistry*, 221, 66-71.
- GINTHER, S., POSTEL, W., WIERING, H. AND GORG, A. (1988). Acid phosphatase typing for breeding nematode-resistant tomatoes by isoelectric focusing with an ultranarrow immobilized pH gradient. *Electrophoresis*, 9, 616-620.
- HANASH, S.M., STRAHLER, J.R., NEEL, J.V., HAILAT, N., MELHEM, R., KEIM, D., ZHU, X.X., WAGNER, D., GAGE, D.A. AND WATSON, J.T. (1991). Highly resolving two-dimensional gels for protein sequencing. *Proceedings of the National Academy of Sciences USA*, 88, 5709-5713.
- HARRINGTON, M.G., COFFMAN, J.A., CALZONE, F.J., HOOD, L.E., BRITTEN, R.J. AND DAVIDSON, E.H. (1992). Complexity of sea urchin embryo nuclear proteins that contain basic domains. *Proceedings of the National Academy of Sciences USA*, 89, 6252-6256.
- HARRINGTON, M.G., LEE, K.H., YUN, M., ZEWERT, T., BAILEY, J.E. AND HOOD, L.E. (1993). Mechanical precision in two-dimensional electrophoresis can improve spot positional reproducibility. *Applied and Theoretical Electrophoresis*, 3, 347-353.
- HELLMAN, U., WERNSTEDT, C., GONZ, J. AND HELDIN, C.-H. (1995). Improvement of an in-gel digestion for the micropreparation of internal protein fragments for amino acid sequencing. *Analytical Biochemistry*, 224, 451-455.
- HENZEL, W.J., BILLECI, T.M., STULTS, J.T., WONG, S.C., GRIMLEY, C. AND WATANABE, C. (1993). Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proceedings of the National Academy of Sciences USA*, 90, 5011-5015.
- HESS, D., COVEY, T.C., WINZ, R., BROWNSEY, R.W. AND AEBERSOLD, R. (1993). Analytical and micropreparative peptide mapping by high performance liquid chromatography/electrospray mass spectrometry of proteins purified by gel electrophoresis. *Protein Science*, 2, 1342-1351.
- HOBOM, U., HOLTHAEVE, T. AND SANDER, C. (1994). Amino acid analysis and protein database compositional search as a rapid and inexpensive method to identify proteins. *Analytical Biochemistry*, 222, 202-209.
- HOCHSTRASSER, D.F. AND MERRIL, C.R. (1988). 'Catalysts' for polyacrylamide gel polymerization and detection of proteins by silver staining. *Applied and Theoretical Electrophoresis*, 1, 35-40.
- HOCHSTRASSER, D.F., PATCHORNIK, A. AND MERRIL, C.R. (1988). Development of polyacrylamide gels that improve the separation of proteins and their detection by silver staining. *Analytical Biochemistry*, 173, 412-423.
- HOCHSTRASSER, A.C., JAMES, R.W., PONETTA, D. AND HOCHSTRASSER, D.F. (1991a). Preparative isoelectrofocusing and high resolution two-dimensional electrophoresis for concentration and purification of proteins. *Applied and Theoretical Electrophoresis*, 1, 333-337.
- HOCHSTRASSER, D.F., APPEL, R.D., VARGAS, R., PERKIER, R., VU RLOD, J.F., RAVIER, F., PASQUALLI, C., FUNK, M., PELLIGRINI, C., MÜLLER, A.F. AND SCHIERRER, J.R. (1991b). A clinical molecular scanner: the Melanie project. *Medical Computing*, 8, 85-91.
- HOCHSTRASSER, D.F., FRUTIGER, S., PAQUET, N., BAIROCH, A., RAVIER, F., PASQUALLI, C., SANCHEZ, J.-C., TISSOT, J.-D., BJELLOVIST, B., VARGAS, R., APPEL, R.D. AND HUGHES, G.J. (1992). Human liver protein map: a reference database established by microsequencing and gel comparison. *Electrophoresis*, 13, 992-1001.
- HOLT, T.G., CHANG, C., LAURENT-WINTER, C., MURAKAMI, T., DAVIES, J.E. AND THOMPSON, C.J. (1992). Global changes in gene expression related to antibiotic synthesis in *Streptomyces hygroscopicus*. *Molecular Microbiology*, 6, 969-980.
- HONORE, B., LEFFERS, H., MADSEN, P. AND CELIS, J.E. (1993). Interferon-gamma up-regulates a unique set of proteins in human keratinocytes. Molecular cloning and expression of the cDNA encoding the RGD-sequence containing protein IGUP 1-5111. *European Journal of Biochemistry*, 218, 421-430.
- HUBERTY, M.C., VATH, J.E., YU, W. AND MARTIN, S.A. (1993). Site-specific carbohydrate

- identification in recombinant proteins using MALD-TOF MS. *Analytical Chemistry*, 65, 2791-2800.
- HUGHES, G.J., FRITIGER, S., PAQUET, N., PASQUALI, C., SANCHEZ, J.-C., TISSOT, J.D., BAIROCH, A., APPEL, R.D. AND HOCHSTRASSER, D.F. (1993). Human liver protein map update 1993. *Electrophoresis*, 14, 1216-1222.
- HUGHES, J.H., MACK, K. AND HAMPARIAN, V.V. (1988). India ink staining of proteins on nylon and hydrophobic membranes. *Analytical Biochemistry*, 173, 18-25.
- JAMES, P., QUADRONI, M., CARAFOLI, E. AND GONNET, G. (1993). Protein identification by mass profile fingerprinting. *Biochemical and Biophysical Research Communications*, 195, 56-64.
- JIL, H., WHITEHEAD, R.H., REID, G.E., MORITZ, F.L., WARD, L.D. AND SIMPSON, R.J. (1994). Two-dimensional electrophoretic analysis of proteins expressed by normal and cancerous human crypts: application of mass spectrometry to peptide-mass fingerprinting. *Electrophoresis*, 15, 391-405.
- JOHNSON, R.S. AND WALSH, K.A. (1992). Sequence analysis of peptide mixtures by automated integration of Edman and mass spectrometric data. *Protein Science*, 1, 1083-1091.
- JOHNSTON, R.F., PICKETT, S.C. AND BARKER, D.L. (1990). Autoradiography using storage phosphor technology. *Electrophoresis*, 11, 355-360.
- JUNGBLIT, P., DZIONARA, M., KLOSE, J. AND WITTMANN-LEIBOLD, B. (1992). Identification of tissue proteins by amino acid analysis after purification by two-dimensional electrophoresis. *Journal of Protein Chemistry*, 11, 603-612.
- JUNGBLIT, P., OTTO, A., ZEINDL-EBERHART, E., PLEIBNER, K.-P., KNECHT, M., REGITZ-ZAGROSEK, V., FLECK, E. AND WITTMANN-LEIBOLD, B. (1994). Protein composition of the human heart: the construction of a myocardial two-dimensional electrophoresis database. *Electrophoresis*, 15, 685-707.
- KOHARA, Y., AKIYAMA, K. AND ISONO, K. (1987). The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. *Cell*, 50, 495-508.
- KLOSE, J. (1975). Protein mapping by combined isoelectric focusing and electrophoresis in mouse tissues. A novel approach to testing for individual point mutations in mammals. *Human Genetics*, 26, 231-243.
- LATHAM, K.E., GARRELS, J.I., CHANG, C. AND SOLTER, D. (1991). Quantitative analysis of protein synthesis in mouse embryos I: extensive re-programming at the one- and two-cell stages. *Development*, 2, 921-932.
- LATHAM, K.E., GARRELS, J.I., CHANG, C. AND SOLTER, D. (1992). Analysis of embryonic mouse development: construction of a high-resolution, two-dimensional gel protein database. *Applied and Theoretical Electrophoresis*, 2, 163-170.
- LATHAM, K.E., GARRELS, J.I. AND SOLTER, D. (1993). Two-dimensional analysis of protein synthesis. *Methods in Enzymology*, 255, 473-489.
- LE MAIRE, M., DESCHAMPS, S., MOLLER, J.V., LE CAER, J.P. AND ROSSIER, J. (1993). Electrospray ionization mass spectrometry from sodium dodecyl sulfate-polyacrylamide gel electrophoresis: application to the topology of the sarcoplasmic reticulum  $\text{Ca}^{2+}$ -ATPase. *Analytical Biochemistry*, 214, 50-57.
- LEMKIN, P.F. AND LESTER, E.P. (1989). Database and search techniques for two-dimensional gel protein data: a comparison of paradigms for exploratory data analysis and prospects for biological modelling. *Electrophoresis*, 10, 122-140.
- LEMKIN, P.F., WU, Y. AND UPTON, K. (1993). An efficient disk-based data structure for rapid searching of quantitative two-dimensional gel databases. *Electrophoresis*, 14, 1341-1350.
- LI, K.W., GERAERTS, W.P., VAN-ELK, R. AND KOOSE, J. (1989). Quantification of proteins in the subnanogram and nanogram range: comparison of the AutoDye, FerriDye, and India ink staining methods. *Analytical Biochemistry*, 182, 44-47.
- LIANG, P. AND PARDEE, A.B. (1992). Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science*, 257, 967-971.
- MASS, M. (1995). Sequence database searching by mass spectrometric data. In *Microcharacterisation of Proteins* (R. Kellner, F. Lottspeich, and H.E. Meyer, Eds), pp 223-245. VCH, Weinheim.

- MANN, M., HOJRUP, P. AND ROEPSTORFF, P. (1993). Use of mass spectrometric molecular weights to identify proteins in sequence databases. *Biological Mass Spectrometry*, **22**, 338-345.
- MANN, M. AND WILM, M. (1994). Error tolerant identification of peptides in sequence databases by peptide sequence tags. *Analytical Chemistry*, **66**, 1390-1399.
- MATSUMURA, P. (1987). Sequence of picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. *Journal of Biological Chemistry*, **262**, 10035-10038.
- MONARD, P.J., BOLTELL, T., GARRELS, J.I. AND LATTER, G.J. (1994). A distributed system for two-dimensional gel analysis. *Computer Applications in the Biosciences*, **10**, 137-143.
- MORTZ, E., VOPAL, O., MANN, M. AND ROEPSTORFF, P. (1994). Identification of proteins in polyacrylamide gels by mass spectrometric peptide mapping combined with database search. *Biological Mass Spectrometry*, **23**, 249-261.
- MURTHY, L.R. AND LOBAL, K. (1991). Measurement of picomoles of phosphoamino acids by high performance liquid chromatography. *Analytical Biochemistry*, **193**, 299-303.
- MYRICK, J.E., LEMKIN, P.F., ROBINSON, M.K. AND LIPTON, K.M. (1993). Comparison of the Biomege Vimage 2000 and the GELLAB-II two-dimensional electrophoresis image analysis systems. *Applied and Theoretical Electrophoresis*, **3**, 335-346.
- NEIDHARDT, F.C., APPLEBY, D.B., SANKAR, P., HUTTON, M.E. AND PHILLIPS, T.A. (1989). Genomically linked cellular protein databases derived from two-dimensional polyacrylamide gel electrophoresis. *Electrophoresis*, **10**, 116-122.
- NIKODEM, V. AND FRESCO, J.R. (1979). Protein fingerprinting by SDS-gel electrophoresis after partial fragmentation with CNBr. *Analytical Biochemistry*, **97**, 382-386.
- NOKIHARA, K., MORITA, N. AND KURIKI, T. (1992). Applications of an automated apparatus for two-dimensional electrophoresis. Model TEP-1, for microsequence analysis of proteins. *Electrophoresis*, **13**, 701-707.
- O'FARRELL, P.H. (1975). High resolution two-dimensional electrophoresis of proteins. *Journal of Biological Chemistry*, **250**, 4007-4021.
- O'FARRELL, P.Z., GOODMAN, H.M. AND O'FARRELL, P.H. (1977). High resolution two-dimensional electrophoresis of basic as well as acidic proteins. *Cell*, **12**, 1133-1142.
- OLIVER *et al.* (1992). The complete DNA sequence of yeast chromosome III. *Nature* **357**, 38-46.
- OLSEN, A.D. AND MILLER, M.J. (1988). Elsie 4: quantitative computer analysis of sets of two-dimensional gel electrophoretograms. *Analytical Biochemistry*, **169**, 49-70.
- ORTIZ, M.L., CALERO, M., FERNANDEZ-PATRON, C., PATRON, C.F., CASTELLANOS, L. AND MENDEZ, E. (1992). Imidazole-SDS-Zn reverse staining of proteins in gels containing or not SDS and microsequence of individual unmodified electroblotted proteins. *FEBS Letters*, **296**, 300-304.
- OSTERGREN, K., ERIKSSON, G. AND BJELLOVIST, B. (1988). The influence of support material used on band sharpness in immobiline gels. *Journal of Biochemical and Biophysical Methods*, **16**, 165-170.
- OU, K., WILKINS, M.R., YAN, J.X., GOOLEY, A.A., FUNG, Y., SHELMAK, D. AND WILLIAMS, K.L. (1995). Improved high-performance liquid chromatography of amino acids derivatised with 9-fluorenylmethyl chloroformate. *Journal of Chromatography* (in press).
- PACKER, N., WILKINS, M.R., GOLAZ, O., LAWSON, M., GOOLEY, A.A., HOCHSTRASSER, D.F., REDMOND, J. AND WILLIAMS, K.L. (1995). Characterisation of human plasma glycoproteins separated by two-dimensional gel electrophoresis. *BioTechnology* (in press).
- PAPPIN, D.J.C., HOJRUP, P. AND BLEASBY, A.J. (1993). Rapid identification of proteins by peptide-mass fingerprinting. *Current Biology*, **3**, 327-332.
- PATTERSON, S.D. (1994). From electrophoretically separated protein to identification: strategies for sequence and mass analysis. *Analytical Biochemistry*, **221**, 1-15.
- PATTERSON, S.D. AND LATTER, G.J. (1993). Evaluation of storage phosphor imaging for quantitative analysis of 2-D gels using the Quest II system. *BioTechniques*, **15**, 1076-1083.
- PISANO, A., REDMOND, J.W., WILLIAMS, K.L. AND GOOLEY, A.A. (1993). Glycosylation sites identified by solid-phase Edman degradation: O-linked glycosylation motifs on human glycoprotein A. *Glycobiology*, **3**, 429-435.
- RABILLOU, T. (1992). A comparison between low background silver diamine and silver nitrate protein stains. *Electrophoresis*, **13**, 429-439.

- RASMUSSEN, H.H., VAN DAMME, J., PIJPE, M., GESSER, B., CELIS, J.E. AND VANDEKERCKHOVE, J. (1992). Microsequences of 145 proteins recorded in the two-dimensional gel protein database of normal human epidermal keratinocytes. *Electrophoresis*, 13, 960-969.
- RASMUSSEN, H.H., MORTZ, E., MANN, M., ROEPSTORFF, P. AND CELIS, J.E. (1993). Identification of transformation sensitive proteins recorded in human two-dimensional gel protein databases by mass-spectrometric peptide mapping alone and in combination with microsequencing. *Electrophoresis*, 15, 406-416.
- RICHARDSON, F.C., HORN, D.M. AND ANDERSON, N.L. (1994). Dose-responses in rat hepatic protein modification and expression following exposure to the rat hepatocarcinogen methapyrilene. *Carcinogenesis*, 15, 325-329.
- RIGHETTI, P.G. (1990). Immobilized pH gradients: theory and methodology. In *Laboratory Techniques in Biochemistry and Molecular Biology* (R.H. Burdon and P.H. van Knippenberg, Eds) Elsevier, Amsterdam.
- RIGHETTI, P.G. AND DRYSDALE, J.W. (1973). *Annals of the New York Academy of Sciences*, 209, 163-186.
- RODRIGUEZ, L.V., GERNSTEN, D.M., RAMAGLI, L.S. AND JOHNSTON, D.A. (1993). Towards stoichiometric silver staining of proteins resolved in complex two-dimensional electrophoresis gels: real-time analysis of pattern development. *Electrophoresis*, 14, 628-637.
- ROSENFELD, J., CAPDEVIELLE, J., GUILLENOT, J.C. AND FERRARA, P. (1992). In-gel digestion of proteins for internal sequence analysis after one- or two-dimensional gel electrophoresis. *Analytical Biochemistry*, 203, 173-179.
- SANCHEZ, J.C., RAVIER, F., PASQUALI, C., FRUTIGER, S., PAQUET, N., BJELLOVIST, B., HOCHSTRASSER, D.F. AND HUGHES, G.J. (1992). Improving the detection of proteins after transfer to polyvinylidene difluoride membranes. *Electrophoresis*, 13, 715-717.
- SANGER, F., COLLSON, A.R., HONG, G.F., HILL, D.F. AND PETERSEN, G.B. (1962). Nucleotide sequence of bacteriophage  $\lambda$  DNA. *Journal of Molecular Biology*, 162, 729-773.
- SCHEELE, G.J. (1975). Two-dimensional analysis of soluble proteins. *Biochemistry*, 250, 5375-5385.
- SHAW, G. (1993). Rapid identification of proteins. *Proceedings of the National Academy of Sciences USA*, 90, 5138-5142.
- SIBBALD, P.R., SOMMERFELDT, H. AND ARGOS, P. (1991). Identification of proteins in sequence databases from amino acid composition. *Analytical Biochemistry*, 198, 330-333.
- SIMPSON, R.J., TSUGITA, A., CELIS, J.E., GARRELS, J.I. AND MEWES, H.W. (1992). Workshop on two-dimensional gel protein databases. *Electrophoresis*, 13, 1055-1061.
- SINHA, P.K., KOTTGEN, E., STOFFLER, M.-M., GIANAZZA, E. AND RIGHETTI, P.G. (1990). Two-dimensional maps in very acidic immobilized pH gradients. *Journal of Biochemical and Biophysical Methods*, 20, 345-352.
- SMITH, D.W. (1994). Introduction. In *Biocomputing: Informatics and Genome Projects* (D.W. Smith, Ed.), pp1-12. Academic Press, San Diego.
- STRUPAT, K., KARAS, M., HILLENKAMP, F., ECKERSKORN, C. AND LUTTSPEICH, F. (1994). Matrix-assisted laser desorption ionization mass spectrometry of proteins electroblotted after polyacrylamide gel electrophoresis. *Analytical Chemistry*, 66, 46-470.
- SUTTON, C.W., PEMBERTON, K.S., COTTRELL, J.S., CORBETT, J.M., WHEELER, C.H., DUNN, M.J. AND PAPPIN, D.J. (1995). Identification of myocardial proteins from two-dimensional gels by peptide mass fingerprinting. *Electrophoresis*, 16, 308-316.
- TOLS, G.I., FAUSNAUGH, J.L., AKINYOYE, O., LACKLAND, H., WINTERCASH, P., VITORICA, F.J. AND STEIN, S. (1989). Amino acid analysis on polyvinylidene difluoride membranes. *Analytical Biochemistry*, 179, 50-55.
- TOVEY, E.R., FORD, S.A. AND BALDO, B.A. (1987). Protein blotting on nitrocellulose: some important aspects of the resolution and detection of antigens in complex extracts. *Journal of Biochemical and Biophysical Methods*, 14, 1-17.
- URWIN, V.E. AND JACKSON, P. (1993). Two-dimensional polyacrylamide gel electrophoresis of proteins labeled with the fluorophore monobromobimane prior to first-dimensional isoelectric focusing: imaging of the fluorescent protein spot patterns using a cooled charge-coupled device. *Analytical Biochemistry*, 209, 57-62.
- VANBOGELEN, R.A., HUTTON, M.E. AND NEIDHARDT, F.C. (1990). Gene-protein database



- of *Escherichia coli*. N-12, edition 3. *Electrophoresis*, 11, 1131-1166.
- VANBOGELEN, R.A. AND NEIDHARDT, F.C. (1991). The gene-protein database of *Escherichia coli*, edition 4. *Electrophoresis*, 12, 955-994.
- VANBOGELEN, R.A., SANKER, F., CLARK, R.L., BOGAN, J.A. AND NEIDHARDT, F.C. (1992). The gene-protein database of *Escherichia coli*, edition 5. *Electrophoresis*, 13, 101-105.
- VANDEKERKHOVE, J., BALW, G., VANCOMPERNOLLE, K., HONORE, B. AND CELIS, J. (1990). Comparative two-dimensional gel analysis and microsequencing identifies gelsolin as one of the most prominent downregulated markers of transformed human fibroblast and epithelial cells. *Journal of Cell Biology*, 111, 95-102.
- VANFLETEREN, J.R., RAYMACKERS, J.G., VAN BUN, S.M. AND MEHUS, L.A. (1992). Peptide mapping and microsequencing of proteins separated by SDS-PAGE after limited *in situ* hydrolysis. *BioTechniques*, 12, 550-557.
- VORM, O. AND MANN, M. (1994). Improved mass accuracy in matrix-assisted laser desorption/ionization time-of-flight mass spectrometry of peptides. *Journal of the American Society for Mass Spectrometry*, 5, 955-958.
- VORM, O., ROEPSTORFF, P. AND MANN, M. (1994). Improved resolution and very high sensitivity in MALDI TOF of matrix surfaces made by fast evaporation. *Analytical Chemistry*, 66, 3281-3287.
- WALLACE, A. AND SALLZ, H.P. (1992a). Ultramicrodetection of proteins in polyacrylamide gels. *Analytical Biochemistry*, 203, 27-34.
- WALLACE, A. AND SALLZ, H.P. (1992b). Beyond silver staining. *Nature*, 357, 606-609.
- WALSH, B.J., GOOLEY, A.A., WILLIAMS, K.L. AND BRETT, S.N. (1995). Identification of macrophage activation associated proteins by two-dimensional electrophoresis and microsequencing. *Journal of Leukocyte Biology*, 57, 507-512.
- WASINGER, V.C., CORDWELL, S.J., POLJAK, A., YAN, J.X., GOOLEY, A.A., WILKINS, M.R., DUNCAN, M., HARRIS, R., WILLIAMS, K.L. AND HUMPHERY-SMITH, I. (1995). Progress with Gene-Product Mapping of the Mollicutes: *Mycoplasma genitalium*. *Electrophoresis*, 16, In Press.
- WEITZHANDLER, M., KADLECEK, D., AVDALOVIC, N., FORTE, J.G., CHOW, D. AND TOWNSEND, R. R. (1993). Monosaccharide and oligosaccharide analysis of proteins transferred to polyvinylidene fluoride membranes after sodium dodecyl sulfate-polyacrylamide gel electrophoresis. *Journal of Biological Chemistry*, 268, 5121-5130.
- WILKINS, M.R., PASQUALI, C., APPEL, R.D., OU, K., GOLAZ, O., SANCHEZ, J.-C., YAN, J.X., GOOLEY, A.A., HUGHES, G., HUMPHERY-SMITH, I., WILLIAMS, K.L. AND HOCHSTRASSER, D.F. (1995). From Proteins to Proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. Submitted.
- WILKINS, M.R., OU, K., APPEL, R.D., GOLAZ, O., PASQUALI, C., YAN, J.X., FARNSWORTH, V., CARTIER, P., HOCHSTRASSER, D.F., WILLIAMS, K.L. AND GOOLEY, A.A. (1996). Rapid protein identification using N-terminal sequence tagging and amino acid analysis (submitted).
- WIRTH, P.J., LUO, L.D., FUJIMOTO, Y., BISGAARD, H.C. AND OLSEN, A.D. (1991). The rat liver epithelial (RLE) cell protein database. *Electrophoresis*, 12, 931-954.
- WIRTH, P.J., LUO, L.D., BENJAMIN, T., HUANG, T.N., OLSEN, A.D. AND PARMALEE, D.C. (1993). The rat liver epithelial (RLE) cell nuclear protein database. *Electrophoresis*, 14, 1199-1215.
- WU, Y., LEMKIN, P.F. AND UPTON, K. (1993). A fast spot segmentation algorithm for two-dimensional gel electrophoresis analysis. *Electrophoresis*, 14, 1351-1356.
- YAMAGUCHI, K. AND ASAKAWA, H. (1988). Preparation of colloidal gold for staining proteins electrotransferred onto nitrocellulose membranes. *Analytical Biochemistry*, 172, 104-107.
- YAMASHITA, K., IDEO, H., OHKURA, T., FUKUSHIMA, K., YUASA, I., OHNO, K. AND TAKESHITA, K. (1993). Sugar chains of serum transferrin from patients with carbohydrate deficient glycoprotein syndrome. Evidence of asparagine-N-linked oligosaccharide transfer deficiency. *Journal of Biological Chemistry*, 268, 5783-5789.
- YATES, J.R. III, SPEICHER, S., GRIFFIN, P.R. AND HUNKAPILLER, T. (1993). Peptide mass maps: a highly informative approach to protein identification. *Analytical Biochemistry*, 214, 397-408.

The

R.M.D.

Therm  
Hamilt.  
Wales  
Private

Introdu

Protein  
dealing  
type. T  
structur  
since th  
by Wa  
1930) &  
possibly  
mechar  
subtilis  
1986; I  
comme  
proteas  
the char  
been in  
regulat  
apparai  
Kalisz  
Barrett  
Give  
surpris  
attracti  
subject  
stabilit  
enzym  
Sherod

• Contes  
Research  
1214-872



**This Page is Inserted by IFW Indexing and Scanning  
Operations and is not part of the Official Record**

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☒ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☒ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** \_\_\_\_\_

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**